Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94*

THOMAS A. HALGREN

Department of Molecular Design and Diversity, Merck Research Laboratories, Rahway, New Jersey 07065

Received 20 March, 1995; accepted 31 August, 1995

ABSTRACT

This article introduces MMFF94, the initial published version of the Merck molecular force field (MMFF). It describes the objectives set for MMFF, the form it takes, and the range of systems to which it applies. This study also outlines the methodology employed in parameterizing MMFF94 and summarizes its performance in reproducing computational and experimental data. Though similar to MM3 in some respects, MMFF94 differs in ways intended to facilitate application to condensed-phase processes in molecular-dynamics simulations. Indeed, MMFF94 seeks to achieve MM3-like accuracy for small molecules in a combined "organic/protein" force field that is equally applicable to proteins and other systems of biological significance. A second distinguishing feature is that the core portion of MMFF94 has primarily been derived from high-quality computational data—ca. 500 molecular structures optimized at the HF/6-31G* level, 475 structures optimized at the MP2/6-31G* level, 380 MP2/6-31G* structures evaluated at a defined approximation to the MP4SDQ/TZP level, and 1450 structures partly derived from MP2/6-31G* geometries and evaluated at the MP2/TZP level. A third distinguishing feature is that MMFF94 has been parameterized for a wide variety of chemical systems of interest to organic and medicial chemists, including many that feature frequently occurring combinations of functional groups for which little, if any, useful experimental data are available. The methodology used in parameterizing MMFF94 represents a fourth distinguishing feature. Rather than using the common "functional group" approach, nearly all MMFF parameters have been determined in a mutually consistent fashion from the full set of available computational data. MMFF94 reproduces the computational data used in its parameterization very well. In addition, MMFF94 reproduces experimental bond lengths (0.014 Å root mean square [rms]), bond angles (1.2° rms), vibrational frequencies (61 cm^{-1}

*This article includes Supplementary Material available from the author upon request or via the Internet at ftp.wiley.com/public/journals/jcc/suppmat/17/490 or http://journals.wiley.com/jcc

Journal of Computational Chemistry, Vol. 17, Nos. 5 & 6, 490–519 (1996)

CCC 0192-8651 / 96 / 050490-30

rms), conformational energies (0.38 kcal/mol rms), and rotational barriers (0.39 kcal/mol rms) very nearly as well as does MM3 for comparable systems. MMFF94 also describes intermolecular interactions in hydrogen-bonded systems in a way that closely parallels that given by the highly regarded OPLS force field.

Introduction

Molecular-mechanics force fields are a crucial component in the armamentarum used by computational and medicinal chemists for what has become known as "rational drug design." Early forms of such force fields go back to the work of Hendrickson[1] in the 1960s. Many would find particularly noteworthy the work of Allinger and coworkers in developing MM1,[2] MM2,[3] and MM3[4]; that of Kollman and coworkers in developing AMBER[5]; that of Jorgensen and coworkers in developing OPLS[6]; that of Karplus and coworkers in developing CHARMM[7]; and that of Lifson and coworkers in developing CVFF.[8] Recent developments, some of which chart important new directions, include the extended CHARMM force field of Momany and Rone[9]; the DREIDING force field of Mayo et al.[10]; the UFF force field of Rappé, et al.[11]; the YETI force field of Vedani and Huhta[12]; the SHAPES and VALBOND force fields of Landis and coworkers[13,14]; the CFF93 force field developed for the Biosym Consortium on Potential Energy Functions by Hagler and coworkers[15,16]; and the MM4 force field of Allinger et al.[17] Like all contemporary force fields, each of the above employs significant physical approximations that limit its accuracy. Moreover, each applies to a different portion of organic/bio-organic chemistry and is derived in a distinctive fashion from a specific selection of data. Given the severity of the approximations that have had to be made, force-field development has been as much an art as a science. One result is that little consensus has been forged either as to what form the force field should take or as to how it should be derived and tested.

One can imagine a different situation—one in which essentially all physically significant effects are incorporated accurately into the force field and in which alternative approximations for specific physical terms can be rigorously tested and validated. What makes such a situation imaginable is the steadily increasing computational power available to computational chemists. Such computational power simultaneously makes it possible to employ more complex and more accurate force fields in molecular simulations and to obtain high-quality computational data against which to determine the form of the force field and on which to base its parameterization. Computational theory has already reached the point at which practical *ab initio* methods routinely give results for small-molecule properties that approach experiment in accuracy[18] while avoiding the large errors that experiment sometimes incurs.[19] Moreover, the requisite computational data can be obtained relatively easily for essentially any system of interest, including many for which no pertinent experimental data are, or are likely to become, available. Arguably, then, a complex, broadly parameterized force field even now can best be derived from computational data.

Efforts to develop improved force fields using computational data and computationally derived insights are already underway in various laboratories. In particular, much has been learned about how to model molecular charge distributions accurately[20] and to incorporate induced-dipole effects arising from molecular polarizability.[21] These electrostatic terms critically affect nonbonded interactions. Research based on the use of computational data obtained from *ab initio* calculations has also been undertaken to better define the form and improve the parameterization of the valence-coordinate terms that depend on bond, angle, and torsional distortions.[16,22] Particularly noteworthy in the latter regard are the novel fits of the empirical potential energy expression to *ab initio* relative energies and first and second derivatives employed by Hagler and coworkers in their derivation of CFF93.[16]

In this series of articles, we report the results of our own initial effort to employ computationally derived information to develop an improved molecular mechanics/dynamics force field. We should note, however, that we have not relied exclusively on computational data. In particular, we have supplemented and extended the range of the core, computationally derived force field, which itself is quite broad, by also parameterizing the force field against a large number of crystallographically determined structures. This combined

effort has led to what we call the Merck Molecular Force Field (MMFF). We call this initial published version "MMFF94." We should note at the outset that MMFF94 still makes significant approximations in its treatment of important physical interactions. Even this version, however, employs computational data of higher quality and broader range than we believe has been utilized in previous efforts. This effort also embodies a particular point of view on what a force field intended for use in bio-organic and pharmaceutical applications should do and on how it should be derived and validated. We expect that growing computational power will soon allow a computationally based approach to be implemented in an even more comprehensive fashion to develop a physically superior force field. In the meantime, we believe the performance and range of applicability of MMFF94 warrant its description and use in computational simulations. To this end, we have deposited the parameters as supplementary material in computer-readable form.[23] Part or all of each parameter file is listed in this or in one of the other articles in this series.[24-27] Moreover, we have collaborated with others to implement MMFF93 in CHARMm[28] and are working to make MMFF94 available in CHARMM,[7] the academic version. In addition, MMFF94 is currently being implemented in the BatchMin module of the MacroModel program suite.[29] We also hope to be able to distribute OPTIMOL,[30] the host molecular-mechanics program for which MMFF94 was developed, through the Quantum Chemistry Program Exchange.[31]

In the next section, we first state the philosophy that underlies the development of MMFF94. We define the form of MMFF94. The fourth section briefly compares the forms of the MMFF94, MM2X, MM2, MM3, and CFF93 force fields. We then define the range of chemical structures for which the computationally derived "core" portion of MMFF has been parameterized and characterize the computational data used. Next, we outline the methodology employed in deriving the force field. We then summarize how MMFF94 performs against computational and experimental data in meeting various structural and energetic tests, and subsequently we describe some elements of its implementation in OPTIMOL, CHARMm, and BatchMin. Finally, we summarize this work and sketch some future directions we believe force-field development will take.

Subsequent articles in this series will complete the description of MMFF94 by more fully defining: (a) the parameterization of the van der Waals (vdW) and electrostatic representation (part II[24]); (b) the parameterization of the valence-coordinate terms that determine molecular geometries and vibrational frequencies (part III[25]); (c) the parameterization of the torsion terms that then determine conformational energies and torsional barriers (part IV[26]); and (d) the further extension of MMFF using a combination of experimetnal data extracted from the Cambridge Crystallographic Database, additional computational data, and carefully calibrated empirical rules (part V[27]). Each of these reports also further characterizes the performance of the new force field in reproducing computational and experimental data. In this introductory article, we summarize MMFF94's performance and address the issues that unify its derivation.

One further clarification needs to be made. This version of MMFF is primarily intended for use in molecular-dynamics simulations rather than in energy-minimization studies. As a practical matter, the principal distinction between these applications concerns MMFF94's treatment of low-energy inversion barriers at resonance-delocalized tricoordinate nitrogen in amides and in such unsaturated amines as vinylamines, anilines, guanines, and nucleic-acid bases. In particular, MMFF94 usually gives nonplanar energy-minimized geometries at nitrogen, even for amides, thereby emulating the nonplanar MP2/6-31G*-optimized geometries used in its parameterization. Yet experimental structures, particularly those determined via crystallographic techniques, tend to show planar or nearly planar geometries that reflect time-averaged atomic positions. When used in molecular-dynamics simulations, MMFF94 produces relatively flat dynamically averaged structures for such species. Many current pharmaceutical applications, however, rely on energy-minimization methods because of limitations in software and computational resources. For use in such studies, we are developing and intend to soon describe a modified version, currently called "MMFF94s," that yields nearly planar energy-minimized geometries for delocalized trigonal nitrogen.[32] The two force fields share most parameters and yield similar, often identical, results for other systems.

Basis and Motivation for Formulation of MMFF94

A molecular mechanics/dynamics force field may reasonably be asked to reproduce accurately any or all of a number of molecular properties,

including the following:

- molecular geometries.
- conformational and stereoisomeric energies.
- torsional barriers and torsion-deformation energies.
- intermolecular-interaction energies.
- intermolecular-interaction geometries.
- vibrational frequencies.
- heats of formation.

Ideally, a single force field would be capable of reproducing these and other molecular properties accurately both in gas-phase and in condensed-phase simulations. Because of their relatively simple construction, however, current force fields necessarily make a variety of compromises. Here we discuss the choices we have made in developing MMFF94.

A pivotal application for MMFF94, from which a number of constraints on its design and implementation follow, is the study of receptor–ligand interactions involving proteins or nucleic acids as receptors and a wide range of chemical structures as ligands. For quantitative study, the force field must be able to describe the ligand and receptor properly in isolation as well as when bound. For these purposes, molecular geometries need to be good, but conformational energies are crucial if the force field is to avoid modeling the wrong conformer of the ligand (or receptor) upon binding or giving an erroneous estimate of the energetic cost of adopting the detailed conformation required for binding. To assess these aspects properly, the force field must be able to locate conformational minima accurately and describe intervening torsional profiles and barriers reasonably well.

At least equally importantly, intermolecular-interaction energies (and, to a lesser extent, geometries) must also be described accurately. In contrast, vibrational frequencies should be reasonably accurate, but spectroscopic precision is unlikely to be required. Thus, fine details of vibrational spectra, such as the splitting of high-frequency modes for bond stretching or angle bending, are unlikely to appreciably affect the *differential* free energy of binding to a macromolecular receptor of one ligand relative to another. Finally, though heats of formation are crucial in some applications, they are not required to understand differences in free energies of binding and are not addressed in MMFF94.

To be routinely and reliably useful in pharmaceutical, bio-organic and chemical applications, MMFF94 would need to be able to handle most organic structural types represented in the *Merck Index*[33] or the *Fine Chemicals Directory*.[34] This broad intended range of application places significant requirements on the data to be used in the parameterization of the force field. We note in this regard that Allinger and coworkers have crafted a series of highly regarded molecular-mechanics force fields based primarily on the meticulous examination and careful selection of good quality experimental data,[2–4,17] and that other force fields such as AMBER[5] and CHARMM[7] have also been parameterized mainly against experimental data. This approach, however, could not be used to derive MMFF94, for two reasons. First, the location, selection, and extraction of good experimental data is a highly time-consuming enterprise and requires a degree of expertise we lack. Second, and more importantly, high-quality experimental data, particularly for conformational and intermolecular-interaction energies, are unavailable for a great many of the chemical structures MMFF94 must handle.

For these reasons, the core portion of MMFF94, on which we focus here, has been derived primarily from *ab initio* data (though experimental data have been liberally employed in its validation). An especially cogent argument for the use of such computational data has recently been offered by Hagler and coworkers.[16a] In a novel and noteworthy departure from previous practice, these workers employed data for molecular dipole moments, relative energies, and Cartesian first and second derivatives obtained from HF/6-31G* calculations to characterize the quantum mechanical energy surface used to derive the QMFF (quantum-mechanical force field) predecessor of CFF93, the Biosym Consortium force field. The approach we have taken in deriving MMF94 is, in part, patterned after theirs. Both, for example, employ the powerful Consortium program PROBE[35] to derive force constants for terms related to bond stretching and angle bending from the information on the curvature of the quantum mechanical surface contained in the HF/6-31G* second derivatives. However, the two approaches also differ in a number of ways that may materially affect their performance in molecular simulations.[24–26]

The derivation of a force field from computational data would be straightforward if we wished to describe only gas-phase systems. However,

while many, and perhaps most, of the processes we wish to model occur in condensed phases, MMFF94 accounts for the effects of molecular polarizability only in a limited way. These effects, for example, cause the dipole moment of water to rise from a gas-phase value of 1.85 D to a mean value of ~2.4 D in aqueous solution.[21d,36] Clearly, a condensed-phase simulation that uses a gas-phase dipole moment for water would seriously underestimate electrostatic interactions and would be expected to yield poor computational properties.[37] Consequently, MMFF94, like OPLS[6] and other current force fields intended for use in condensed-phase simulations, employs *effect pair potentials*[37] that reflect, in an averaged sense, the enhancement of the charge distribution due to molecular polarizability.

Especially careful attention must be given to the partitioning between electrostatic, van der Waals (vdW), and torsional interactions.[38] Our approach begins by choosing the vdW representation as previously defined[39] and the electrostatic representation from fits to scaled (enhanced by 10%)[24] HF/6-31G*[40] molecular dipole moments. To properly describe hydrogen-bonding interactions, we then adjust key vdW and electrostatic parameters to better fit scaled intermolecular-interaction energies and geometries obtained from HF/6-31G* calculations.[41] Last, we derive the torsion terms to fit the *ab initio* gas-phase conformational data. Conveniently, the HF/6-31G* level of theory consistently overestimates gas-phase dipole moments for organic compounds.[6b,18a] For water, it gives a calculated dipole moment of 2.20 D,[42] or 2.42 D after 10% enhancement, close to the previously cited mean value of ~2.4 D found in aqueous solution. This use of scaled HF/6-31G* interaction energies and geometries allows MMFF to be parameterized in a straightforward manner that seeks to ensure that a proper balance between solvent–solvent, solvent–solute, and solute–solute interactions is achieved. The quality of this balance is crucial for accurately describing aqueous solvation and the energetics of host/guest binding in aqueous solution. We have also explored the use of higher level *ab initio* calculations,[24] but have not found an alternative approach that appears preferable.

As noted in the Introduction, one further choice we made in developing MMFF94 was to derive a force field explicitly intended for use in molecular-dynamics simulations. A modified version more suitable for use in energy minimization studies (MMFF94s) is also being developed.[32]

Form of the Merck Molecular Force Field

The MMFF94 energy expression can be written as:

$$E_{MMFF} = \sum EB_{ij} + \sum EA_{ijk} + \sum EBA_{ijk} + \sum EOOP_{ijk;l} + \sum ET_{ijkl} + \sum EvdW_{ij} + \sum EQ_{ij} \quad (1)$$

where the seven constituent terms are defined as shown below. In each case, the cited numerical constant is such that the deformation or interaction energy is expressed in kilocalories per mole when distances and angles are measured in angstroms and in degrees, respectively.

In the notation that follows, we adopt the convention that a specific atom involved in a force-field interaction is designated by $i, j, k, \ldots$ and that the corresponding numeric MMFF atom type is designated by $I, J, K, \ldots$ This notation makes explicit, for example, that the force constant kb_{IJ} and the reference bond length r^0_{IJ} for the i–j bond in eq. (2) depend on the associated MMFF atom types I and J, whereas the bond distance, r_{ij}, depends on the atomic coordinates. In certain instances, the parameters depend only on the atomic species for atoms $i, j, k, \ldots$; in such cases, we still use capital letters, but explicitly note the actual dependence in the text.

BOND STRETCHING

MMFF94 employs the quartic function:

$$EB_{ij} = 143.9325\frac{kb_{IJ}}{2}\Delta r_{ij}^2 \times \left(1 + cs\,\Delta r_{ij} + 7/12\,cs^2\,\Delta r_{ij}^2\right) \quad (2)$$

where kb_{IJ} is the force constant (md/Å), $\Delta r_{ij} = r_{ij} - r^0_{IJ}$ is the difference (Å) between actual and reference bond lengths, and $cs = -2$ Å^{-1} is the "cubic-stretch" constant. This function corresponds to an expansion through fourth order of a Morse function with an "alpha" of 2 Å^{-1}.[43] Results published in a recent high-level *ab initio* study[44] show this value for alpha to be a representative one. Special sets of reference bond lengths

and force constants are employed for "conjugated single bonds," such as those found in butadiene and biphenyl, as well as for certain other single bonds between *sp*- or sp^2-hybridized atoms.[25]

ANGLE BENDING

MMFF94 normally uses the cubic expansion:

$$\mathrm{EA}_{ijk} = 0.043844\frac{\mathrm{k}a_{IJK}}{2}\Delta\vartheta_{ijk}^2(1 + cb\,\Delta\vartheta_{ijk}) \quad (3)$$

where $\mathrm{k}a_{IJK}$ is the force constant (md Å/rad^2), $\Delta\vartheta_{ijk} = \vartheta_{ijk} - \vartheta_{IJK}^0$ is the difference between actual and reference bond angles (degrees), and $cb = -0.007$ deg^{-1} (or, more precisely, -0.4 rad^{-1}) is the "cubic-bend" constant. Special sets of parameters are used for angles that involve delocalized single bonds and/or occur in small rings.[25] For linear or near-linear bond angles, MMFF94 employs the well-behaved form used in DREIDING[10] and UFF[11]:

$$\mathrm{EA}_{ijk} = 143.9325\mathrm{k}a_{IJK}(1 + \cos\vartheta_{ijk}) \quad (4)$$

STRETCH-BEND INTERACTIONS

MMFF94 employs the form:

$$\mathrm{EBA}_{ijk} = 2.51210(\mathrm{k}ba_{IJK}\,\Delta r_{ij} + \mathrm{k}ba_{KJI}\,\Delta r_{kj})\,\Delta\vartheta_{ijk} \quad (5)$$

where $\mathrm{k}ba_{IJK}$ and $\mathrm{k}ba_{KJI}$ are force constants (md/rad) that couple the i–j and k–j stretches to the i–j–k bend, and Δr and $\Delta\vartheta$ are as defined above. Stretch–bend interactions are omitted when eq. (4) is used for bond angles.

OUT-OF-PLANE BENDING AT TRICOORDINATE CENTERS

MMFF94 uses the form:

$$\mathrm{EOOP}_{ijk;l} = 0.043844\frac{\mathrm{k}oop_{IJK:L}}{2}\chi_{ijk;l}^2 \quad (6)$$

where $\mathrm{k}oop_{IJK:L}$ is the force constant (md Å/rad^2) and $\chi_{ijk;l}$ is the Wilson angle[45] (degrees) between the bond j–l and the plane i–j–k. The three angles that arise at a given center, j, are all assigned the same $\mathrm{k}oop_{IJK:L}$ force constant; the "in-plane" angles use "normal" bond angles and are described by eq. (3). For trigonal nonplanar centers, this formulation allows angle-bending reference values that average less than 120° to be used to make the center pyramidal; the out-of-plane term can then be employed to improve the fit to the inversion barrier.

TORSION INTERACTIONS

MMFF94 uses the threefold representation employed in MM2[3] and MM3,[4] where Φ is the i–j–k–l torsion angle:

$$\mathrm{ET}_{ijkl} = 0.5(V_1(1 + \cos\Phi) + V_2(1 - \cos 2\Phi) + V_3(1 + \cos 3\Phi)) \quad (7)$$

The constants V_1, V_2, and V_3 depend on the atom types I, J, K, and L for atoms i, j, k, and l, where i–j, j–k, and k–l are bonded pairs. Torsion interactions within four-membered rings and saturated five-membered rings[26] are given special torsion constants, as are interactions in which either the central or a wing bond is a single bond between two atoms that are capable of participating in multiple or aromatic bonds.[26] The former situation occurs, for example, in biphenyl, butadiene, and styrene.

VAN DER WAALS INTERACTIONS

MMFF employs the recently developed "Buffered-14-7" form[39]; the terminology derives from the formal 14th and 7th power dependencies for the repulsive and attractive terms that would be obtained if the R_{IJ}^* "buffering constants" in the denominators were deleted. The form of the potential is shown in eq. (8):

$$\mathrm{E}_{\mathrm{vdW}_{ij}} = \varepsilon_{IJ}\left(\frac{1.07R_{IJ}^*}{R_{ij} + 0.07R_{IJ}^*}\right)^7\left(\frac{1.12R_{IJ}^{*7}}{R_{ij}^7 + 0.12R_{IJ}^{*7}} - 2\right) \quad (8)$$

This form is used in conjunction with an expression that relates the minimum-energy separation R_{II}^* to the atomic polarizability α_I [eq. (9)], with specially formulated combination rules [eqs. (10) and (11)], and with a Slater–Kirkwood expression for the well depth ε_{IJ} [eq. (12)]:

$$R_{II}^* = A_I\alpha_I^{1/4} \quad (9)$$

$$R_{IJ}^* = 0.5(R_{II}^* + R_{JJ}^*)\left(1 + 0.2\left(1 - \exp(-12\gamma_{IJ}^2)\right)\right) \quad (10)$$

$$\gamma_{IJ} = (R_{II}^* - R_{JJ}^*)/(R_{II}^* + R_{JJ}^*) \quad (11)$$

$$\varepsilon_{IJ} = \frac{181.16G_IG_J\alpha_I\alpha_J}{(\alpha_I/N_I)^{1/2} + (\alpha_J/N_J)^{1/2}}\frac{1}{R_{IJ}^{*6}} \quad (12)$$

As described elsewhere,[24] modified values of R^*_{IJ} and ε_{IJ} are used to describe hydrogen-bonding interactions. Van der Waals and electrostatic interactions are included whenever atoms i and j belong to separate domains or are separated by three or more chemical bonds; 1,4-vdW interactions are not differentially scaled in MMFF94.

ELECTROSTATIC INTERACTIONS

MMFF94 uses the buffered coulombic form:

$$EQ_{ij} = 332.0716 q_i q_j / \left(D(R_{ij} + \delta)^n \right) \quad (13)$$

where q_i and q_j are partial atomic charges, R_{ij} is the internuclear separation in Å, $\delta = 0.05$ Å is the "electrostatic buffering" constant, and D is the "dielectric constant." Normally, the exponent n is taken as 1, though use of a distance-dependent dielectric constant ($n = 2$) is also supported. In MMFF94, 1,4-electrostatic interactions are scaled by a factor of 0.75.[26] The distance buffering, where $\delta > 0$, prevents infinite attractive electrostatic energies from overwhelming the finite repulsive vdW interaction contained in eq. (8) as oppositely charged atomic centers coalesce.

The partial atomic charges q_i used in eq. (13) are constructed from initial full or fractional "formal atomic charges" q_I^0 (usually zero, but, e.g., $+1/3$ for guanidinium nitrogen) by adding contributions from bond charge increments ω_{KI} that describe the polarity of the bonds to atom i from attached atoms k. Specifically, MMFF94 computes q_i as

$$q_i = q_I^0 + \sum \omega_{KI} \quad (14)$$

where $\omega_{KI} = -\omega_{IK}$. The procedure used to assign the q_I^0 is specified in part V of our study.[27]

Comparison of MMFF94's Functional Form to MM2, MM2X, MM3, and CFF93

MMFF94 closely resembles MM2 and MM3, as well as MM2X, our previous generation force field,[46] in functional form. For bond stretching, MMFF94 and MM3 each use a quartic expansion in which the cubic and quartic force constants are related to the quadratic force constants in a predetermined way. Each thereby avoids the "cubic stretch" catastrophe, in which progressive elongation of a chemical bond eventually drives the MM2 or MM2X energy to negative infinity. This catastrophe is circumvented in MM2X's implementation at the cost of additional complexity in the computer code that might prove troublesome in molecular-dynamics applications. MMFF94 and MM3 employ anharmonic angle bending, an intrinsically better representation than the simple quadratic form used in MM2 and MM2X, though MMFF truncates its representation at the cubic term.[47] Moreover, trigonal centers are handled differently in MMFF94 to allow out-of-plane terms to be used for certain centers that have nonplanar equilibrium geometries.[25] Centers having linear idealized bond angles are also handled differently, through eq. (4). The same forms for stretch–bend and torsion interaction are used in all four force fields. MMFF94 currently omits MM3's bond–torsion and bend–bend terms; bond–torsion terms may be included later in certain cases, as may angle–torsion terms. MM3 also includes electronegativity-related adjustments to reference bond lengths that MMFF94 omits but that in certain cases can be significant.[25,27]

The most important differences between MMFF94 and the MM2, MM2X, and MM3 force fields arise in the description of nonbonded interactions. In particular, MMFF94 uses a "buffered" expression for vdW interactions and employs novel combination rules for the vdW parameters. Unlike MM2X,[46] MMFF94 properly treats intramolecular and intermolecular electrostatic interactions in the same manner. Moreover, MMFF94 normally utilizes a unit dielectric constant, thereby allowing the force field to be applied without modification to condensed-phase simulations employing explicit solvent. In addition, like AMBER,[5] CHARMM,[7] CVFF,[8], CFF93,[16] and most other force fields used in molecular-dynamics simulations, MMFF94 describes hydrogen-bonding interactions as being essentially electrostatic in nature, whereas MM2 (1987 parameters and later) and MM3 obtain up to 6 kcal/mol of stabilization energy[48] from an attractive Exp-6 term.

CFF93 and MMFF94 both use a quartic expansion for bond stretching, treat stretch–bend interactions in the same way, and employ equivalent representations for torsion interactions.[16] Both, in addition, define partial atomic charges in terms of bond charge increments, describe electrostatic interactions solely in terms of charge–charge interactions (i.e., avoid special "hydrogen-bond" terms[5,7]), and use novel vdW combination rules[39,49] in conjunction with a vdW potential (Lennard–Jones 9-6 or Buf-14-7) that differs from the more commonly used Lennard–Jones 12-6

form. Minor differences include MMFF94's use of a cubic rather than quartic expansion for angle bending and its use of a special form [eq. (4)] to describe "linear bond angles." One major difference is that CFF93 includes many more "cross terms." These terms allow CFF93 to describe certain elements of geometry more accurately, an example being the elongation of a conjugated single bond by as much as 0.1 Å upon bond rotation. The additional cross terms also allow CFF93 to reproduce vibrational spectra more accurately than can MMFF94. As noted earlier, however, MMFF94 does not seek to predict vibrational spectra to high accuracy but rather to describe conformational and intermolecular interaction energies as well as possible. With respect to these latter considerations, major differences in parameterization that may materially affect performance arise from differences in the data and methodology used.[24-26]

Survey of Systems for Which MMFF94 is Parameterized

To define the range of organic and bio-organic systems covered in the core parameterization of MMFF94, the set of compounds and molecular conformations for which optimized MP2/6-31G* geometries have been employed are listed in Table I. Each structure is identified by a five-character conformational index of the form "*XYNMc*," where "XY" defines the compound class (e.g., AM for amides and related compounds), "*NM*" specifies the compound number within the class, and "*c*" identifies the conformation. The conformational designations "*a*" through "*i*" correspond to equilibrium conformers; "*j*" through "*z*" indicate conformations optimized while holding a particular torsion angle fixed, except that "*t*" and sometimes "*s*" usually denote a symmetry-determined conformational transition state. To characterize the conformation, the compound name is followed, where appropriate, by a brief description of the geometry.

Among "monofunctional" chemical families, MMFF94 has been parameterized for alkanes, alkenes, alcohols, phenols, ethers, aldehydes, ketones, ketals, acetals, hemiketals, hemiacetals, amines, amides, peptide analogs, ureas, imides, carboxylic acids, esters, carboxylate anions, ammonium cations, thiols, mercaptans, disulfides, halides (chlorides and fluorides), imines, iminium

TABLE I.
Conformers Considered in Parameterizing Core MMFF94.[a]

Amides and peptide analogs

AM01a —formamide
AM01t —formamide, N planar
AM02a —*N*-methylformamide, cis
AM02b —*N*-methylformamide, trans
AM02j —*N*-methylformamide, trans, h—c—n—c = 60°
AM02k —*N*-methylformamide, trans, h—c—n—c = 30°
AM02l —*N*-methylformamide, trans, h—c—n—c = 0°
AM02s —*N*-methylformamide, ~anti ts, h—n—c=o = 115°
AM02t —*N*-methylformamide, ~anti ts, h—n—c=o = 120°
AM02u —*N*-methylformamide, ~anti ts, h—n—c=o = 125°
AM02v —*N*-methylformamide, ~anti ts, h—n—c=o = 130°
AM02w —*N*-methylformamide, ~syn ts, h—n—c=o = 55°
AM02x —*N*-methylformamide, ~syn ts, h—n—c=o = 60°
AM02y —*N*-methylformamide, ~syn ts, h—n—c=o = 65°
AM02z —*N*-methylformamide, ~syn ts, h—n—c=o = 70°
AM03a —acetamide
AM03t —acetamide, N planar
AM04a —*N*-methylacetamide, trans
AM04b —*N*-methylactamide, cis
AM04j —*N*-methylacetamide, trans, h—c—c=o = 0°
AM04k —*N*-methylacetamide, trans, h—c—c=o = 30°
AM04l —*N*-methylacetamide, trans, h—c—c=o = 60°
AM04m —*N*-methylacetamide, cis, h—c—c=o = 0°
AM04s —*N*-methylacetamide, cis, N planar
AM04t —*N*-methylacetamide, trans, N planar
AM05a —*N*,*N*-dimethylformamide
AM06a —urea, puckered
AM06t —urea, planar
AM07a —*N*-formylformamide, both o=c—n—h cis
AM07b —*N*-formylformamide, o=c—n—h cis, trans
AM08a —formylglycinamide
AM09a —glycine dipeptide analog, C7
AM09b —glycine dipeptide analog, C5

(*Continues on next page*)

TABLE I.
(continued)

AM09s — glycine dipeptide analog, C7, N planar
AM09t — glycine dipeptide analog, C5, N planar
AM10a — alanine dipeptide analog, C7eq
AM10b — alanine dipeptide analog, C5
AM10c — alanine dipeptide analog, C7ax
AM10d — alanine dipeptide analog, α'
AM10e — alanine dipeptide analog, β_2
AM10f — alanine dipeptide analog, α_L
AM11a — propionamide, c — c — c — n anti
AM12a — *N*-ethylformamide, c — c — n — c gauche
AM12a — *N*-ethylformamide, c — c — n — c = 180°
AM13a — *N*-OH,*N*-methylacetamide, o — n — c= o trans
AM13b — *N*-OH,*N*-methylacetamide, o — n — c= o cis
AM13s — *N*-OH NMA, o — n — c= o trans, N planar
AM13t — *N*-OH NMA, o — n — c= o cis, N planar
AM14a — *N*-OH,N — Et acetamide, o — n — c= o trans, c — c — n — o gauche
AM14b — *N*-OH,N — Et acetamide, o — n — c= o trans, c — c — n — o trans
AM14c — *N*-OH,N — Et acetamide, o — n — c= o cis, c — c — n — c(=o) skew
AM14d — *N*-OH, N — Et acetamide, o — n — c= o cis, c — c — n — c(=o) gauche
AM15a — *N*-OH, N-Me propionamide, o — n — c= o trans, c — c — c= o cis
AM15b — *N*-OH, N-Me propionamide, o — n — c= o trans, c — c — c= o skew
AM15c — *N*-OH, N-Me propionamide, o — n — c= o cis, c — c — c= o cis
AM15d — *N*-OH, N-Me propionamide, o — n — c= o cis, c — c — c= o skew
AM16a — glycine dipeptide, C7
AM16b — glycine dipeptide, C5
AM17a — alanine dipeptide, C7eq
AM17b — alanine dipeptide, C5
AM17c — alanine dipeptide, C7ax
AM17d — alanine dipeptide, α'
AM17e — alanine dipeptide, β_2
AM17f — alanine dipeptide, α_L

TABLE I.
(continued)

Carboxylate anions
AN01a — formate anion
AN02a — acetate ion
AN03a — propionate anion
AN04a — propenoate anion

Aromatic and heteroaromatic compounds
AR01a — benzene
AR02a — pyridine
AR03a — pyrimidine
AR04a — pyridazine
AR05a — 1,3,5-triazine
AR06a — pyrrole
AR07a — furan
AR08a — thiophene
AR09a — imidazole
AR10a — pyrazole
AR11a — 1,2,4-triazole
AR12a — 1,2,3,4-tetrazole (N1 tautomer)
AR13a — 1,2,3,5-tetrazole (N2 tautomer)
AR14a — oxazole
AR15a — isoxazole
AR16a — 1,3,4-oxadiazole
AR17a — 1,2,4-oxadiazole
AR18a — thiazole
AR19a — isothiazole
AR20a — 1,3,4-thiadiazole
AR21a — pyridine N-oxide
AR22a — toluene
AR23a — ethylbenzene, c — c — c — c skew
AR23t — theylbenzene, c — c — c — c = 0°
AR24a — *N*-ethylpyrrole, c — n — ch_2 — ch_3 ca. 90°
AR25a — indole

Carboxylic acids
CA01a — methanoic acid, o= c — o — h cis
CA01b — methanoic acid, o= c — o — h trans
CA02a — ethanoic acid, o= c — o — h cis
CA02b — ethanoic acid, o= c — o — h trans
CA03a — propanoic acid, c — c — c= o cis
CA03b — propanoic acid, c — c — c= o skew
CA04a — glyoxalic acid, o= c — c= o trans, o= c — o — h cis
CA04b — glyoxalic acid, o= c — c= o trans, o= c — o — h trans (h-bond)
CA05a — glycolic acid, o= c — c — o cis (h-bond)
CA05b — glycolic acid, o= c — c — o skew (h-bond)
CA06a — benzoic acid
CA07a — propenoic acid, c= c — c= o trans
CA07b — propenoic acid, c= c — c= o cis
CA08a — oxalic acid, o= c — c= o trans, both o= c — o — h trans
CA08b — oxalic acid, o= c — c= o trans, both o= c — o — h cis

(*Continues on next page*)

TABLE I.
(continued)

CA08c — oxalic acid, o=c—c=o trans, one o=c—o—h trans
CA09a — pyruvic acid, o=c—c=o, o=c—o—h trans (h-bond)
CA09b — pyruvic acid, o=c—c=o, o=c—o—h cis

Carboxylic acid esters

CE01a — methyl formate, o=c—o—c cis
CE01b — methyl formate, o=c—o—c trans
CE01j — methyl formate, o=c—o—c trans, h—c—o—c = 180°
CE01k — methyl formate, o=c—o—c trans, h—c—o—c = 150°
CE01l — methyl formate, o=c—o—c trans, h—c—o—c = 120°
CE02a — methyl acetate, o=c—o—c trans
CE02b — methyl acetate, o=c—o—c cis
CE05a — vinyl formate, o=c—o—c cis, c=c—o—c trans
CE05b — vinyl formate, o=c—o—c cis, c=c—o—c cis
CE06a — ethyl formate, o=c—o—c cis, c—o—c—c anti
CE06b — ethyl formate, o=c—o—c cis, c—o—c—c gauche
CE07a — isopropyl formate, c—c—o—c = g, a
CE07b — isopropyl formate, c—c—o—c = g, g
CE08a — phenyl acetate
CE08b — phenyl acetate, c—c(=o)—o—c ~90°
CE09a — propiolactone
CE10a — methyl glycolate, o=c—c—o cis (h-bond)
CE10b — methyl glycolate, o=c—c—o skew (h-bond)

Conjugated systems

CJ01a — 1,3-butadiene, gauche
CJ01b — 1,3-butadiene, s-trans
CJ01t — 1,3-butadiene, c=c—c=c = 0°
CJ02a — 2-methyl-1,3-butadiene, gauche
CJ02b — 2-methyl-1,3-butadiene, s-trans
CJ03a — 2-methyl-but-1-ene-3-one, c=c—c=o cis
CJ03b — 2-methyl-but-1-ene-3-one, c=c—c=o trans
CJ04a — 2-methylpropenamide, c=c—c=o cis
CJ04b — 2-methylpropenamide, c=c—c—o skew
CJ05a — propenamide, c=c—c=o cis
CJ05b — propenamide, c=c—c=o skew
CJ06a — but-1-ene-3-one, c=c—c=o cis
CJ06b — but-1-ene-3-one, c=c—c=o trans
CJ07a — acrolein, c=c—c=o cis
CJ07b — acrolein, c=c—c=o trans
CJ08a — 2-methylpropenal, c=c—c=o cis
CJ08b — 2-methylpropenal, c=c—c=o trans
CJ09a — 2-methylpropenoic acid, c=c—c=o trans
CJ09b — 2-methylpropenoic acid, c=c—c=o cis

TABLE I.
(continued)

CJ10a — acetophenone
CJ11a — styrene
CJ12a — 2-phenylpropene
CJ12j — 2-phenylpropene, framework planar
CJ13a — 1,3-pentadiene, s-trans, c—c=c—c trans
CJ13b — 1,3-pentadiene, gauche, c—c=c—c trans
CJ13c — 1,3-pentadiene, s-trans, c—c=c—c cis
CJ14a — cyclopentadiene

Aldehydes and ketones

CO01a — formaldehyde
CO02a — acetaldehyde
CO03a — propionaldehyde, c—c—c=o cis
CO03b — propionaldehyde, c—c—c=o
CO04a — acetone
CO05a — butanone c—c—c—o = 0°
CO05b — butanone c—c—c—o skew
CO05j — butanone, c—c—c=o = 0°
CO05k — butanone, c—c—c=o = 30°
CO05l — butanone, c—c—c=o = 60°
CO06m — butanone, c—c—c=o = 90°
CO05n — butanone, c—c—c=o = 120°
CO05o — butanone, c—c—c=o = 150°
CO05p — butanone, c—c—c=o = 180°
CO06a — methyl isopropyl ketone, o=c—c(ch_3)$_2$—h trans
CO06b — methyl isopropyl ketone, o=c—c(ch_3)$_2$—h cis
CO07a — butyraldehyde, c—c—c—c anti
CO07b — butyraldehyde, c—c—c—c gauche
CO08a — but-3-enal c=c—c—c skew, c—c—c=o cis
CO08b — but-3-enal c=c—c—c skew−, c—c—c=o skew+
CO08c — but-3-enal c=c—c—c skew+, c—c—c=o skew+
CO09a — 3-methyl-but-3-enal, c=c—c—c skew, c—c—c=o cis
CO09b — 3-methyl-but-3-enal, c=c—c—c skew, c—c—c=o skew
CO10j — isobutyraldehyde, h—c(=o)—c—h = 0°
CO10k — isobutyraldehyde, h—c(=o)—c—h = 30°
CO10l — isobutyraldehyde, h—c(=o)—c—h = 60°
CO10m — isobutyraldehyde, h—c(=o)—c—h = 90°
CO10n — isobutyraldehyde, h—c(=o)—c—h = 120°
CO10o — isobutyraldehyde, h—c(=o)—c—h = 150°
CO10p — isobutyraldehyde, h—c(=o)—c—h = 180°
CO11a — cyclobutanone
CO11t — cyclobutanone, planar

(*Continues on next page*)

TABLE I. (continued)

CO12a — 2-formylpropanal, o — c — c — c(= o) anti
CO12b — 2-formylpropanal, o — c — c — c(=o) gauche
CO13a — 4-oxobutanal, o= c — c — c, c — c — c= o cis, c — c — c — c trans
CO13b — 4-oxobutanal, o= c — c — c, c — c — c= o cis, c — c — c — c gauche
CO14a — 2,3-butanedione, c — c — c — c trans
CO14t — 2,3-butanedione, c — c — c — c cis

Halides

HL01a — fluoromethane
HL02a — difluoromethane
HL03a — 1,1-difluorethane
HL04a — 1,2-difluoroethane, f — c — c — f anti
HL04b — 1,2-difluoroethane, f — c — c — f gauche
HL05a — 1,2-dichloroethane, cl — c — c — cl anti
HL05b — 1,2-dichloroethane, cl — c — c — cl gauche
HL06a — 1,1,1-trifluoroethane
HL07a — 1,1,1-trichloroethane
HL08a — chlorocyclobutane
HL08j — chlorocyclobutane, planar
HL09a — fluoropropane, c — c — c — f anti
HL09b — fluoropropane, c — c — c — f gauche
HL10a — chloropropane, c — c — c — cl anti
HL10b — chloropropane, c — c — c — cl gauche

Imines, guanadines, and amidines

IM01a — formamidine, h — n= c — n cis, N puckered
IM01b — formamidine, h — n= c — n anti, N puckered
IM01t — formamidine, h — n= c — n cis, N planar
IM02a — *N*-methylformaldehydeimine, h — c — n= c cis
IM02t — *N*-methylformaldehydeimine, h — c — n= c = 180°
IM03a — formaldehydeimine
IM04a — *N*-methylformamidine, n — c= n — c cis, N puckered
IM04b — *N*-methylformamidine, n — c= n — c trans, N puckered
IM04t — *N*-methylformamidine, n — c= n — c cis, N planar
IM05a — guanidine, N puckered
IM05t — guanidine, planar
IM06a — N_2-methylguanidine, N puckered
IM06t — N_2-methylguanidine, N planar
IM07a — butadiene Schiff base, c= c — c= n s-trans, h — n= c — c cis
IM07b — butadiene Schiff base, c= c — c= n s-cis, h — n= c — c trans

TABLE I. (continued)

Ketals, acetals, and hemiacetals

KT02a — 2-methoxytetrahydropyran, equatorial
KT02b — 2-methoxytetrahydropyran, axial, me — o — c — c anti
KT03a — 2,4 dioxapentane, c — o — c — o g+, o — c — o — c g+
KT03b — 2,4 dioxapentane, c — o — c — o g, o — c — o — c a
KT04a — 2,5-dimethyl-1,3-dioxane (5-equatorial)
KT04b — 2,5-dimethyl-1,3-dioxane (5-axial)
KT05a — methoxymethanol, c — o — c — o g+, o — c — o — h g+
KT05b — methoxymethanol, c — o — c — o g+, o — c — o — h g −
KT05c — methoxymethanol, c — o — c — o g, o — c — o — h a

Cations

NC01a — ammonium cation
NC02a — *N*-methylamine cation
NC03a — *N*-ethylamine cation
NC03t — *N*-ethylamine cation, h — n — c — c = 0°
NC04a — *N*,*N*-dimethylamine cation
NC05a — *N*-propylamine cation, c — c — c — n gauche
NC05b — *N*-propylamine cation, c — c — c — n anti
NC06a — guanidine cation
NC07a — ethylguanidine cation, c — c — n= c anti
NC07b — ethylguanidine cation, c — c — n= c gauche
NC08a — formamidine cation
NC09a — methylguanidine cation
NC10a — *N*-methylformaldehydeimine cation
NC11a — *N*-methylformamidine cation, c — n — c= n cis
NC11b — *N*-methylformamidine cation, c — n — c= n trans
NC12a — imidazolium cation
NC13a — formaldehydeimine cation
NC14a — *t*-butylamine cation
OC01a — hydronium ion

Amines

NH01a — methylamine
NH02a — propylamine, c — c — c — n anti
NH02b — propylamine, c — c — c — n gauche
NH03a — isopropylamine, C1 (c — h gauche to n −lp)
NH03b — isopropylamine, Cs (c — h anti to n — lp)
NH03j — isopropylamine, h — c — n — h = 120°
NH03k — isopropylamine, h — c — n — h = 150°
NH03l — isopropylamine, h — c — n — h = 180°
NH03m — isopropylamine, h — c — n — h = 210°
NH03n — isopropylamine, h — c — n — h = 240°
NH03p — isopropylamine, h — c — n — h = 270°
NH03p — isopropylamine, h — c — n — h = 300°

(*Continues on next page*)

TABLE I.
(continued)

NH04a — cyclohexylamine, equatorial
NH04b — cyclohexylamine, axial
NH05a — dimethylamine
NH06a — azetidine, n — h equatorial
NH06j — azetidine, ring planar
NH07a — piperidine, n — h equatorial
NH07b — piperidine, n — h axial
NH08a — trimethylamine
NH09a — *N*-methylpiperidine, equatorial
NH09b — *N*-methylpiperidine, axial
NH10a — ammonia
NH10t — ammonia, planar
NH11a — ethylamine, c — c — n — lp gauche
NH11b — ethylamine, c — c — n — lp anti
NH12a — *t*-butylamine
NH13a — vinylamine
NH14a — aniline, N puckered
NH14t — aniline, planar
NH15a — pyrrolidine, n — h equatorial
NH15j — pyrrolidine, ring planar
NH16a — 3-aminopropene, c= c — c — n skew, c — c — n — lp gauche
NH16b — 3-aminopropene, c= c — c — n cis, c — c — n — lp gauche
NH16c — 3-aminopropene, c= c — c — n skew, c — c — n — lp anti
NH17a — 2-me,3-aminopropene, c= c — c — n skew, c — c — n — lp gauche
NH17b — 2-me,3-aminopropene, c= c — c — n cis, c — c — n — lp gauche
NH18a — ethylenediamine, n — c — c — n anti, c — c — n — lp g+ , g+
NH18b — ethylenediamine, n — c — c — n g+ , c — c — n — lp g+ , g+
NH19a — *N*-methylaniline, N puckered
NH19t — *N*-methylaniline, N planar
NH20a — methylethylamine *N*-oxide, c — n — c — c anti
NH20b — methylethylamine *N*-oxide, c — n — c — c gauche
NH21a — methylethylhydroxylamine, c — n — c — c anti
NH21b — methylethylhydroxylamine, c — n — c — c gauche
NH22a — ethylamine *N*-oxide, o — n — c — c gauche
NH22b — ethylamine *N*-oxide, o — n — c — c anti
NH23a — ethylhydroxylamine, o — n — c — c gauche
NH23b — ethylhydroxylamine, o — n — c — c anti

Alcohols

OH01a — methanol
OH02a — ethanol, c — c — o — h gauche
OH02b — ethanol, c — c — o — h anti

TABLE I.
(continued)

OH02j — ethanol, c — c — o — h = 0°
OH02k — ethanol, c — c — o — h = 30°
OH02l — ethanol, c — c — o — h = 60°
OH02m — ethanol, c — c — o — h = 90°
OH02n — ethanol, c — c — o — h = 120°
OH02o — ethanol, c — c — o — h = 150°
OH02p — ethanol, c — c — o — h = 180°
OH03a — *n*-propanol, c — c — c — o a, c — c — o — h g
OH03b — *n*-propanol, c — c — c — o g− , c — c — o — h g+
OH03c — *n*-propanol, c — c — c — o g+ , c — c — o — h g+
OH03d — *n*-propanol, c — c — c — o a, c — c — o — h a
OH03e — *n*-propanol, c — c — c — o g, c — c — o — h a
OH04a — isopropanol, h — c — o — h gauche
OH04b — isopropanol, h — c — o — h anti
OH04j — isopropanol, h — c — o — h = 0°
OH04k — isopropanol, h — c — o — h = 30°
OH04l — isopropanol, h — c — o — h = 60°
OH04m — isopropanol, h — c — o — h = 90°
OH04n — isopropanol, h — c — o — h = 120°
OH04o — isopropanol, h — c — o — h = 150°
OH04p — isopropanol, h — c — o — h = 180°
OH05a — *t*-butanol
OH06a — cyclopentanol, equatorial Cs
OH06b — cyclopentanol, axial Cs
OH06c — cyclopentanol, equatorial C1
OH06d — cyclopentanol, axial C1
OH06j — cyclopentanol, Cs, ring planar
OH07a — cyclohexanol, equatorial Cs
OH07b — cyclohexanol, axial Cs
OH07c — cyclohexanol, equatorial C1
OH07d — cyclohexanol, axial C1
OH08a — phenol
OH09a — water
OH10a — vinyl alcohol, c= c — o — h trans
OH10b — vinyl alcohol, c= c — o — h skew
OH11a — benzyl alcohol
OH11b — benzyl alcohol, h — o — c — c anti
OH12a — propen-3-ol, c= c — c — o skew, c — c — o — h a
OH12b — propen-3-ol, c= c — c — o cis, c — c — o — h a
OH12c — propen-3-ol, c= c — c — o skew, c — c — o — h g
OH13a — 2-me-propen-3-ol, c= c — c — o s, c — c — o — h a
OH13b — 2-me-propen-3-ol, c= c — c — o c, c — c — o — h a
OH14a — sec-butanol, ga / ag[b]
OH14b — sec-butanol, ga / ga
OH14c — sec-butanol, ga / gg

(*Continues on next page*)

TABLE I.
(continued)

OH14d — sec-butanol, ag / ag
OH14e — sec-butanol, ag / ga
OH14f — sec-butanol, ag / gg
OH14g — sec-butanol, gg / ag
OH14h — sec-butanol, gg / ga
OH14i — sec-butaone, gg / gg
OH14r — sec-butanol, cm / ag, approx ts
OH14s — sec-butanol, cm / gg, approx ts
OH14t — sec-butanol, ga / cm, approx ts
OH14u — sec-butanol, ga / mp, approx ts
OH14v — sec-butanol, ga / pc, approx ts
OH14w — sec-butanol, mp / ag, approx ts
OH14x — sec-butanol, mp / gg, approx ts
OH14y — sec-butanol, pc / ag, approx ts
OH14z — sec-butanol, pc / gg, approx ts
OH15a — 1,2-ethanediol, h — o — c — c a, o — c — c — o a, c — c — o — h a
OH15b — 1,2-ethanediol, h — o — c — c g − , o — c — c — o g+ , c — c — o — h a
OH15c — 1,2-ethanediol, h — o — c — c g − , o — c — c — o g+ , c — c — o — h g +
OH15d — 1,2-ethanediol, h — o — c — c g+ , o — c — c — o g − , c — c — o — h g+

Ethers

OR01a — methyl ethyl ether, c — o — c — c anti
OR01b — methyl ethyl ether, c — o — c — c gauche
OR02a — methyl ethyl ether, c= c — o — c cis
OR02b — methyl ethyl ether, c= c — o — c skew
OR03a — diethyl ether, c — c — o — c anti, c — o — c — c anti
OR03b — diethyl ether, c — c — o — c anti, c — o — c — c gauche
OR04a — methoxycyclohexane, equatorial Cs
OR04b — methoxycyclohexane, axial C1
OR04c — methoxycyclohexane, equatorial C1
OR05a — oxetane, C2
OR05t — oxetane, planar
OR06a — dimethyl ether
OR07a — tetrahydrofuran, C2
OR07t — tetrahydrofuran, ring planar
OR11a — dioxolane, C2
OR11t — dioxolane, ring planar
OR13a — methyl isopropyl ether, h — c — o — ch_3 gauche
OR13b — methyl isopropyl ether, h — c — o — ch_3 anti
OR14a — methyl phenyl ether, c — o — c — c cis
OR14j — methyl phenyl ether, c — o — c — c = 90°

Alkanes

RA01a — methane
RA02a — ethane, staggered
RA02t — ethane, eclipsed

TABLE I.
(continued)

RA03a — propane
RA04a — butane, c — c — c — c anti
RA04b — butane, c — c — c — c gauche
RA04t — butane, c — c — c — c = 0°
RA05a — isobutane
RA06a — cyclobutane
RA06t — cyclobutane, ring planar
RA07a — cyclopentane, half-chair C2
RA07t — cyclopentane, ring planar
RA08a — cyclohexane, chair
RA08b — cyclohexane, twist-boat C2
RA10a — methylcyclohexane, equatorial
RA10b — methylcyclohexane, axial
RA11a — neopentane
RA12a — 2,3-dimethylbutane, h — c2 — c3 — h gauche
RA12b — 2,3-dimethylbutane, h — c2 — c3 — h anti
RA13a — cyclopropane
RA14a — cyclooctane, crown D4d
RA14b — cyclooctane, boat-chair Cs
RA14c — cyclooctane, twist-boat-chair C2
RA14d — cyclooctane, S4
RA15a — methylcyclobutane, equatorial
RA15b — methylcyclobutane, axial
RA15j — methylcyclobutane, ring planar
RA16a — cyclononane, [144] C2
RA16b — cyclononane, [333] D3
RA16c — cyclononane, [225] C2
RA16d — cyclononane, [234] C1
RA16e — cyclononane, [9a] C1 (used in validating MMFF94)

Alkenes

RE01a — ethylene
RE02a — propene
RE03a — 1-butene, c= c — c — c cis
RE03b — 1-butene, c= c — c — c skew
RE04a — 1-pentene, c — c — c — c anti
RE04b — 1-pentene, c — c — c — c gauche
RE05a — 2-methyl-1-butene, c= c — c — c skew
RE05b — 2-methyl-1-butene, c= c — c — c cis
RE06a — isobutene
RE07a — 1,4-pentadiene, c= c — c — c s+ , s −
RE07b — 1,4-pentadiene, c= c — c — c s − , s −
RE08a — trans-2-butene
RE08b — cis-2-butene
RE09a — cyclobutene
RE10a — trans-2-pentene, c — c — c= c skew
RE10b — cis-2-pentene, c — c — c= c skew
RE11a — 1,4-cyclohexadiene

Thiols, sulfides, and disulfides

SR01a — hydrogen sulfide
SR02a — ethanethiol, c — c — s — h gauche

(*Continues on next page*)

TABLE I.
(continued)

SR02b — ethanethiol, c — c — s — h anti
SR03a — dimethyl sulfide
SR04a — ethyl methyl disulfide, c — c — s — s gauche
SR04b — ethyl methyl disulfide, c — c — s — s anti
SR05a — methyl hydrogen disulfide
SR06a — dimethyl disulfide
SR07a — thiophenol
SR07t — thiophenol, planar
SR08a — methyl phenyl sulfide, c — c — s — c ca. 90°
SR08t — methyl phenyl sulfide, c — c — s — c = 0°
SR09a — 1-propanethiol, s — c — c — c anti, h — s — c — c gauche
SR09b — 1-propanethiol, s — c — c — c g − , h — s — c — c g+
SR09c — 1-propanethiol, s — c — c — c gauche, h — s — c — c anti
SR10a — methyl hydrogen sulfide
SR11a — 1,2-ethanedithiol, all anti
SR11b — 1,2-ethanedithiol, h — s — c — c anti, anti; s — c — c — s gauche
SR11c — 1,2-ethanedithiol, h — s — c — c anti, g+ ; s — c — c — s g −
SR12a — methyl propyl sulfide, c — s — c — c gauche, s — c — c — c anti
SR12b — methyl propyl sulfide, c — s — c — c g − , s — c — c — c g −
SR12c — methyl propyl sulfide, c — s — c — c anti, s — c — c — c gauche
SR12f — methyl propyl sulfide, c — s — c — c anti, s — c — c — c anti
SR13a — thiomethanol

[a]For brevity, the conformational abbreviations a, g, t, c, and s are sometimes used for anti, gauche, trans, cis, and skew, respectively. These designations correspond approximately to torsion angles of 180°, 60°, 180°, 0°, and 120°. Where appropriate, the abbreviations "g+" and "g−" or "s+" and "s−" are used to indicate the relative signs of gauche or skew angles.

[b]For the OH14 conformers (sec-butanol), in the conformational designations "wx/yz," "w" indicates the conformation of the c — c — c — o angle, "x" that of the c — c — c — c angle, "y" that of the h — o — c — ch_2 angle, and "z" that of the h — o — c — ch_3 angle; the designations "m" and "p" conote angles of approximately −120° and 120°, respectively.

cations, amine *N*-oxides, hydroxylamines, hydroxamic acids, amidines, guanidines, amidinium cations, guanidinium cations, imadazolium cations, aromatic hydrocarbons, and heteroaromatic compounds (cf. Table II). As can be inferred from the listing in Table I, the structural coverage is quite broad for some of these chemical families but is limited for others. Many of the bifunctional compounds included in the parameterization are unsaturated analogs of families listed above, that is, conjugated alkenes and aromatic hydrocarbons (e.g., styrenes); α,β-unsaturated variants of amides, imines, aldehydes, ketones, carboxylic acids, esters, and carboxylate anions; vinylic ethers, alcohols, amines, and esters; and allylic aldehydes, ketones, amines, and alcohols. Other bifunctional compounds include: β-ketoacids; β-hydroxyesters; dicarboxylic acids; 1,2-diols, 1,2-diamines, and 1,2-dithiols; and nonconjugated dienes. A limited selection of alkanes, amines, ketones, halides, esters and ethers containing four- or five-membered rings has also been studied. Compounds containing SO_2 and oxyphosphorus groups have been treated as a part of the extension of the core parameterization described in part V.[27]

The number of chemical families treated in the core parameterization of MMFF94 is therefore large—certainly over 20—and many, though by no means all, combinations of functional groups of interest to medicinal and chemical industry chemists have been treated. Nevertheless, an increase of severalfold in the number of core MMFF94 parameters would probably be needed to allow the core force field to handle virtually any organic compound of pharamaceutical interest. To make MMFF94 as useful as possible, we have extended the core force field: (i) by parameterizing MMFF against a large set of experimental structures extracted from the Cambridge Structural Database and against additional computational data; and (ii) by implementing a well-defined set of default-parameter assignments and carefully calibrated empirical rules for parameters not defined by either the structural data or the additional computational data.[27] Some indication of the resultant range of chemical structures covered by MMFF94 can be gleaned from an examination of Table III, which characterizes the current MMFF atom types.

We hope to broaden the core, computationally derived, parameterization of MMFF in future work. Even now, however, the current set of core parameters is, we believe, significantly broader than is provided in other specifically parameterized force fields.[3,4,16] In addition, we believe that the breadth and quality of the extended parameterization compares favorably, for organic compounds, with that provided by other force fields that employ generic

TABLE II.
Classes of Compounds Included in MMFF94's Core Parameterization

- Alkanes, alkenes, aromatic hydrocarbons, conjugated alkenes and aromatics, nonconjugated alkenes
- Five- and six-membered heteroaromatics
- Alcohols, phenols, ethers, 1,2-diols, vinylic alcohols and ethers, allylic alcohols
- Amines, imines, vinylic amines, allylic amines α,β-unsaturated imines, amidines, guanidines, 1,2-diamines
- Hydroxyl amines, amine *N*-oxides
- Amides, dipeptides, ureas, imides, α,β-unsaturated amides, hydroxamic acids
- Aldehydes, ketones, α,β-unsaturated aldehydes and ketones, allylic aldehydes and ketones
- Ketals, acetals, hemiketals, hemiacetals
- Carboxylic acids and esters, vinylic esters, α,β-unsaturated acids and esters, dicarboxylic acids.
- β-ketoacids, β-hydroxyesters
- Thiols, sulfides, disulfides, 1,2-dithiols
- Halides (chlorides and fluorides)
- Amine, imine, amidine, guanidine, pyridine, and imidazole cations
- Carboxylate anions, α,β-unsaturated carboxylate anions
- Various saturated and unsaturated four- and five-membered ring systems

parameters or empirical rules to achieve broad nominal coverage.[10,11]

Computational Data Used in Parameterizing MMFF94

The computational data employed in parameterizing the core force field fall into five main categories[50]:

1. Calculations at the HF/6-31G* level[40] for ~500 HF/6-31G*-optimized geometries. These calculations, carried out using Gaussian 88,[51] Gaussian 90,[52] or Gaussian 92,[53] covered nearly all of the molecular structures and conformations listed in Table I and also included ~70 hydrogen-bonded dimers used in the parameterization of intermolecular interactions.[24]
2. Calculations with full geometry optimization at the MP2/6-31G* level for ~360 equilibrium conformers. This level of theory has been shown to give geometries for standard organic functional groups that rival experiment in accuracy.[18,54] Theoretical geometries are particularly suitable for use in force-field parameterization because they do not entail the assumptions and artificial restrictions sometimes made in deriving experimental geometries,[18] do not require sometimes ill-defined corrections for effects of thermal motion,[18] and are unlikely to manifest the large errors to which experimental determinations occasionally are subject.[19]
3. Calculations for ~380 MP2/6-31G*-optimized geometries carried out at the MP2/TZP level using triple-zeta plus polarization basis sets, and at the MP2 and MP4SDQ levels using modified 6-31G* basis sets. As described in part IV,[26] the MP2/TZP calculations and the MP3 and MP4SDQ corrections obtained using the modified 6-31G* basis set were combined to yield composite energies that we refer to as "MP4SDQ/TZP" energies, where the quotation marks indicate an approximation to full MP4SDQ/TZP results.
4. Single-point MP2/TZP calculations carried out at ~1450 torsionally incremented geometries, derived from MP2/6-31G* geometries and partially optimized using refined but not yet final MMFF94 parameters.
5. Very large basis set calculations on intermolecular interactions in nonpolar systems obtained using highly correlated wavefunctions.

The use of these data in the derivation of MMFF94 is described in what follows.

Methodology Used in Parameterizing MMFF94

The MMFF94 energy expression presented in eq. (1) contains seven terms. For the five valence-coordinate terms, MMFF94 employs quadratic force constants for bond stretching, angle bending, stretch–bend interaction, and out-of-plane bend-

ing; reference values for bond stretching and angle bending; and one or more of the V_1, V_2, and V_3 constants for torsion interactions. Grouping the V_n terms together, MMFF94 therefore utilizes seven classes of valence-coordinate parameters. MMFF94 also employs bond–charge increments ω_{KI} in eq. (14) and atomic polarizabilities α_I in eqs. (9) and (12) to generate quantities used in evaluating nonbonded interactions. In all, then, MMFF94 employs nine classes of force-field parameters. This section outlines the approaches used to derive each such class of parameters, and specifies how the individual approaches were combined to yield a mutually consistent set of parameters. Full details are given elsewhere.[24-27]

NONBONDED VAN DER WALLS AND ELECTROSTATIC PARAMETERS

As noted earlier, representative values for the atomic polarizabilities α_I and for the derived minimum-energy separations R^*_{II} for nonhydrogen atoms have previously been described.[39] A preliminary listing of the associated MMFF atom types has also been given;[39] the current set is specified in Table III. For aliphatic hydrogens, the α and R^* parameters were determined by fitting to high-quality *ab initio* data on intermolecular interactions for the methane[55] and hydrogen dimers. The vdW parameters for the polar hydrogens in water were determined by requiring that the water dimer be described in geometric and energetic terms similar to those found in successfully employed water models. This vdW representation was then transferred to other types of polar hydrogen atoms. Initial values for the bond-charge-increment parameters ω_{KI} in eq. (14) were obtained by employing the Biosym Consortium program PROBE[35] to fit the x, y, and z components of the molecular dipole moments to quantities computed at the HF/6-31G* level and scaled, as discussed in part II,[24] by a factor of 1.10. To make the fit well determined, partial atomic charges for polar hydrogens and other key terminal atoms (and hence the associated ω_{KI} parameters) were fixed at values representative of ESP-fit (electrostatic potential fit)[56] or Mulliken charges obtained from the *ab initio* calculations. Crucially, final vdW and electrostatic parameters for atom types involved in hydrogen-bonding interactions were obtained by adjusting the initial values to fit appropriately scaled[41] interaction energies and hydrogen-bonding geometries computed at the HF/6-31G* level.[24]

TABLE III.
MMFF94 Symbolic and Numeric Atom Types

Atom type		
Symbolic	Numeric	Definition [coordination number][a] {formal charge}[b]
CR	1	Alkyl carbon [4]
C=C	2	Vinylic carbon [3]
CSP2	2	Generic sp^2 carbon [3]
C=O	3	Generic carbonyl carbon [3]
C=N	3	Imine-type carbon [3]
CGD	3	Guanidine carbon [3]
C=OR	3	Ketone or aldehyde carbonyl carbon [3]
C=ON	3	Amide carbonyl carbon [3]
COO	3	Carboxylic acid or ester carbonyl carbon [3]
COON	3	Carbamate carbonyl carbon [3]
COOO	3	Carbonic acid or ester carbonyl carbon [3]
C=OS	3	Thioester carbonyl carbon, double bonded to O [3]
C=S	3	Thioester carbon, double bonded to S [3]
C=SN	3	Thioamide carbon, double bonded to S [3]
CSO2	3	Carbon in $>C=SO_2$ [3]
CS=O	3	Sulfinyl carbon in $>C=S=O$ [3]
CSS	3	Thiocarboxylic acid or ester carbon [3]
C=P	3	Carbon doubly bonded to P [3]
CSP	4	Acetylenic carbon [2]
=C=	4	Allenic carbon [2]
HC	5	Hydrogen attached to carbon [1]
HSI	5	Hydrogen attached to silicon [1]
—O—	6	Generic divalent oxygen [2]
OR	6	Ether oxygen [2]
OC=O	6	Carboxylic acid or ester oxygen [2]
OC=C	6	Enolic or phenolic oxygen [2]
OC=N	6	Oxygen in —O—C=N— moiety [2]
OC=S	6	Divalent oxygen in thioacid or ester [2]
ONO2	6	Divalent nitrate "ether" oxygen [2]
ON=O	6	Divalent nitrate "ether" oxygen [2]
OSO3	6	Divalent oxygen in sulfate group [2]
OSO2	6	Divalent oxygen in sulfite group [2]

(*Continues on next page*)

TABLE III. (continued)

Atom type		
Symbolic	Numeric	Definition [coordination number][a] {formal charge}[b]
OSO	6	One of two divalent oxygens attached to sulfur [2]
OS=O	6	Divalent oxygen in R(RO)S = O [2]
—OS	6	Other divalent oxygen attached to sulfur [2]
OPO3	6	Divalent oxygen in phosphate group [2]
OPO2	6	Divalent oxygen in phosphite group [2]
OPO	6	Divalent oxygen, one of two oxygens attached to P [2]
—OP	6	Other divalent oxygen attached to phosphorus [2]
O=C	7	Generic carbonyl oxygen [1]
O=CN	7	Carbonyl oxygen in amides [1]
O=CR	7	Carbonyl oxygen in aldehydes and ketones [1]
O=CO	7	Carbonyl oxygen in acids and esters [1]
O=N	7	Nitroso oxygen [1]
O=S	7	Doubly bonded sulfoxide oxygen [1]
O=S=	7	O=S on sulfur doubly bonded to, e.g., C [1]
NR	8	Amine nitrogen [3]
N=C	9	Imine nitrogen [2]
N=N	9	Azo-group nitrogen [2]
NC=O	10	Amide nitrogen [3]
NC=S	10	Thioamide nitrogen [3]
NN=C	10	Nitrogen in N—N=C moiety with deloc. lp [3]
NN=N	10	Nitrogen in N—N = N moiety with deloc. lp [3]
F	11	Fluorine [1]
Cl	12	Chlorine [1]
Br	13	Bromine [1]
I	14	Iodine [1]
S	15	Thiol, sulfide, or disulfide sulfur [2]
S=C	16	Sulfur doubly bonded to carbon [1]
S=O	17	Sulfoxide sulfur [3]
> S=N	17	Tricoordinate sulfur doubly bonded to N [3]
SO2	18	Sulfone sulfur [4]
SO2N	18	Sulfonamide sulfur [4]
SO3	18	Sulfonate group sulfur [4]
SO4	18	Sulfate group sulfur [4]

TABLE III. (continued)

Atom type		
Symbolic	Numeric	Definition [coordination number][a] {formal charge}[b]
=SO2	18	Sulfone sulfur, doubly bonded to carbon
SNO	18	Sulfur in nitrogen analog of a sulfone
SI	19	Silicon [4]
CR4R	20	Aliphatic carbon in 4-membered ring [4]
HOR	21	Hydroxyl hydrogen in alcohols [1]
HO	21	Generic hydroxyl hydrogen [1]
CR3R	22	Aliphatic carbon in 3-membered ring [4]
HNR	23	Generic hydrogen on sp^3 nitrogen, e.g., in amines [1]
HPYL	23	Hydrogen on nitrogen in pyrrole [1]
H3N	23	Hydrogen in ammonia [1]
HNOX	23	Hydrogen on N in a N-oxide
HOCO	24	Hydroxyl hydrogen in carboxylic acids [1]
HOP	24	Hydroxyl hydrogen in H—O—P moiety [1]
PO4	25	Phosphate group phosphorus [4]
PO3	25	Phosphorus with 3 attached oxygens [4]
PO2	25	Phosphorus with 2 attached oxygens [4]
PO	25	Phosphine oxide phosphorus [4]
PTET	25	General tetracoordinate phosphorus [4]
P	26	Phosphorus in phosphines [3]
HN=C	27	Hydrogen on imine nitrogen [1]
HN=N	27	Hydrogen on azo nitrogen [1]
HNCO	28	Hydrogen on amide nitrogen [1]
HNCS	28	Hydrogen on thioamide nitrogen [1]
HNCC	28	Hydrogen on enamine nitrogen [1]
HNCN	28	Hydrogen in H—N—C=N moiety [1]
HNNC	28	Hydrogen in H—N—N=C moiety [1]
HNNN	28	Hydrogen in H—N—N = N moiety [1]

(*Continues on next page*)

TABLE III. (continued)

Atom type		
Symbolic	Numeric	Definition [coordination number][a] {formal charge}[b]
HNSO	28	Hydrogen on NSO, NSO2, or NSO3 nitrogen [1]
HNC%	28	Hydrogen on N triply bonded to C [1]
HSP2	28	Generic hydrogen on sp^2 nitrogen [1]
HOCC	29	Enolic or phenolic hydroxyl hydrogen [1]
HOCN	29	Hydroxyl hydrogen in HO—C=N moiety [1]
CR4E	30	Olefinic carbon in 4-membered ring [3]
HOH	31	Hydroxyl hydrogen in water [1]
O2CM	32	Oxygen in carboxylate group [1] {−1/2}
ONX	32	Oxygen in N-oxides [1]
O=N	32	Oxygen in nitroso group [1]
O2N	32	Oxygen in nitro group [1]
O2NO	32	Nitro-group oxygen in nitrate [1]
O3N	32	Nitrate anion oxygen [1] {−1/3}
O—S	32	Single terminal O on tetra-coordinate sulfur [1]
O2S	32	One of 2 terminal O's on sulfur [1] {variable}[c]
O3S	32	One of 3 terminal O's on sulfur [1] {variable}[c]
O4S	32	Terminal O in sulfate anion [1] {−1/2}
OSMS	32	Terminal oxygen in thiosulfinate anion [1] {−1/2}
OP	32	Oxygen in phosphine oxide [1]
O2P	32	One of 2 terminal O's on P [1] {variable}[c]
O3P	32	One of 3 terminal O's on P [1] {variable}[c]
O4P	32	One of 4 terminal O's on P [1] {variable}[c]
O4Cl	32	Oxygen in perchlorate anion [1] {−1/4}
HOS	33	Hydrogen on oxygen attached to sulfur [1]
NR+	34	Quaternary nitrogen [4] {1}
OM	35	Oxide oxygen on sp^3 carbon [1] {−1}
OM2	35	Oxide oxygen on sp^2 carbon [1] {−1}
HNR+	36	Hydrogen on quaternary nitrogen [1]

TABLE III. (continued)

Atom type		
Symbolic	Numeric	Definition [coordination number][a] {formal charge}[b]
HIM+	36	Hydrogen on imidazolium nitrogen [1]
HPD+	36	Hydrogen on pyridinium nitrogen [1]
HNN+	36	Hydrogen on amidinium nitrogen [1]
HNC+	36	Hydrogen on protonated imine nitrogen [1]
HGD+	36	Hydrogen on guanidinium nitrogen [1]
CB	37	Aromatic carbon, e.g., in benzene [3]
NPYD	38	Aromatic nitrogen with σ lone pair [2]
NPYL	39	Aromatic 5-ring nitrogen with π lone pair [2]
NC=C	40	Enamine or aniline nitrogen, deloc. lp [3]
NC=N	40	Nitrogen in N—C=N with deloc. lp [3]
NC=N	40	Nitrogen in N—C = P with deloc. lp [3]
NC%C	40	Nitrogen attached to C—C triple bond [3]
CO2M	41	Carbon in carboxylate anion [3]
CS2M	41	Carbon in thiocarboxylate anion [3]
NSP	42	Triply bonded nitrogen [1]
NSO2	43	Sulfonamide nitrogen [3]
NSO3	43	Sulfonamide nitrogen [3]
NC%N	43	Nitrogen attached to cyano group [3]
STHI	44	Aromatic 5-ring sulfur with π lone pair [2]
NO2	45	Nitrogen in nitro group [3]
NO3	45	Nitrogen in nitrate group [3]
N=O	46	Nitrogen in nitroso group [2]
NAZT	47	Terminal nitrogen in azido or diazo group [1]
NSO	48	Divalent nitrogen replacing monovalent O in SO_2 group [2]
O+	49	Oxonium oxygen [3] {1}
HO+	50	Hydrogen on oxonium oxygen [1]
O=+	51	Oxenium oxygen [2] {1}
HO=+	52	Hydrogen on oxenium oxygen [1]
=N=	53	Central nitrogen in C=N = N or N=N=N [2]
N+=C	54	Iminium nitrogen [3] {1}

(*Continues on next page*)

TABLE III. (continued)

Atom type		
Symbolic	Numeric	Definition [coordination number][a] {formal charge}[b]
N+=N	54	Positively charged nitrogen doubly bonded to N [3] {1}
NCN+	55	Either nitrogen in $N^+=C-N$: [3] {1 / 2}
NGD+	56	Guanidinium nitrogen [3] {1 / 3}
CGD+	57	Guanidinium carbon [3]
CNN+	57	Carbon in $+N=C-N$: resonance structures [3]
NPD+	58	Aromatic nitrogen in pyridinium [3] {1}
OFUR	59	Aromatic 5-ring oxygen with π lone pair [2]
C%−	60	Isonitrile carbon [1]
NR%	61	Isonitrile nitrogen [2]
NM	62	Anionic divalent nitrogen [2] {−1}
C5A	63	Aromatic 5-ring C, α to N:, O:, or S: [3]
C5B	64	Aromatic 5-ring C, β to N:, O:, or S: [3]
N5A	65	Aromatic 5-ring N, α to N:, O:, or S: [2]
N5B	66	Aromatic 5-ring N, β to N:, O:, or S: [2]
N2OX	67	sp^2-hybridized N-oxide nitrogen [3]
N3OX	68	sp^3-hybridized N-oxide nitrogen [4]
NPOX	69	Pyridinium N-oxide nitrogen [3]
OH2	70	Oxygen in water [2]
HS	71	Hydrogen attached to sulfur [1]
HS=N	71	Hydrogen attached to $>S=$ sulfur doubly bonded to N [1]
HP	71	Hydrogen attached to phosphorus [1]
S—P	72	Terminal sulfur bonded to P [1]
SM	72	Anionic terminal sulfur [1] {−1}
SSMO	72	Terminal sulfur in thiosulfinate group [1] {−1 / 2}
SO2M	73	Sulfur in anionic sulfinate group [3]
SSOM	73	Tricoordinate sulfur in anionic thiosulfinate group [3]
=S=O	74	Sulfinyl sulfur, e.g., in C=S=O
—P=C	75	Phosphorus doubly bonded to C [3]

TABLE III. (continued)

Atom type		
Symbolic	Numeric	Definition [coordination number][a] {formal charge}[b]
N5M	76	Nitrogen in 5-ring aromatic anion [2] {variable}[c]
CLO4	77	Perchlorate anion chlorine [4]
C5	78	General carbon in 5-membered heteroaromatic ring [3]
N5	79	General nitrogen in 5-membered heteroaromatic ring [2]
CIM+	80	Aromatic carbon between N's in imidazolium [3]
NIM+	81	Aromatic nitrogen in imidazolium [3] {1 / 2}
N5A+	81	Positive nitrogen in 5-ring alpha position [3] {1}
N5B+	81	Positive nitrogen in 5-ring alpha position [3] {1}
N5+	81	Positive nitrogen in other 5-ring position [3] {1}
N5AX	82	N-oxide nitrogen in 5-ring alpha position [3]
N5BX	82	N-oxide nitrogen in 5-ring beta position [3]
N5OX	82	N-oxide nitrogen in other 5-ring position [3]
FE+2	87	Dipositive iron cation [0] {2}
FE+3	88	Tripositive iron cation [0] {3}
F−	89	Floride anion [0] {−1}
CL−	90	Chloride anion [0] {−1}
BR−	91	Bromide anion [0] {−1}
LI+	92	Lithium cation [0] {1}
NA+	93	Sodium cation [0] {1}
K+	94	Potassium cation [0] {1}
ZN+2	95	Dipositive zinc cation [0] {2}
CA+2	96	Dipositive calcium cation [0] {2}
CU+1	97	Monopositive copper cation [0] {1}
CU+2	98	Dipositive copper cation [0] {2}
MG+2	99	Dipositive magnesium cation [0] {2}

[a] Number of attached atoms.

[b] Initial full or fractional charge, from which final MMFF94 partial atomic charges are obtained by adding contributions arising from the relative polarity of bonds involving attached atoms.

[c] The formal charge is determined by dividing the net ionic charge on the SO_x or PO_x group among the equivalent terminal oxygens.

GEOMETRIC PARAMETERS

The reference bond lengths, r_{IJ}^0, and bond angles, ϑ_{IJK}^0, that appear in eqs. (2) and (3) were determined as follows. Given a trial set of MMFF parameters, optimized MMFF geometries for the molecules used in the parameterization were obtained from, and then systematically compared to, the reference *ab initio* geometries.[57] For each distinct type of bond or angle (as determined by the MMFF atom types and the "bond-type" or "angle-type" index[25]), the average signed deviation between the MMFF and the *ab initio* bond lengths or angles was then determined and was used to adjust the trial reference value. The iterative procedure was initiated by setting the trial reference values equal to the average of the actual bond lengths or angles observed in the *ab initio* structures. As discussed in part III,[25] this approach had to be modified slightly to determine reference angles in small-ring compounds. This procedure was applied both to the MP2/6-31G*-optimized structures and to a similar set of HF/6-31G*-optimized geometries (ca. 350 structures in each case); the reference bond lengths and angles derived from fitting to the HF/6-31G* geometries were used in the fits to the HF/6-31G* first and second derivatives described in the following subsection.

QUADRATIC FORCE CONSTANTS

Force constants for bond stretching, angle bending, stretch-bend interaction, and out-of-plane bending were determined by using the Biosym Consortium program PROBE[35] to fit a slightly modified version of the MMFF94 energy expression to the Cartesian first and second derivatives of the HF/6-31G* energy. The principal modification consisted in replacing MMFF94's Buf-14-7 and buffered electrostatic terms by Lennard–Jones 10-6 and simple coulombic [$\delta = 0$ in eq. (13)] terms. As earlier work has suggested that valence–coordinate force constants are not strongly affected even by the *neglect* of nonbonded interactions,[43] the substitution of comparable terms seems unlikely to have had an appreciable effect on the derived force constants. In these fits, only the quadratic force constants were optimized; parameters of all other classes were held constant. Finally, the HF/6-31G*-derived quadratic force constants were modified for use in MMFF94 by applying scaling factors chosen to optimize the fit of MMFF to experimental vibrational frequencies. Further details are given in part III.[25]

TORSION PARAMETERS

The V_1, V_2, and V_3 parameters in eq. (7) were derived from fits to conformational energies using TORFIT.[58] These fits used "penalty function" restraints in connection with a "build-up" protocol in which all but certain twofold parameters initially were given zero values. The *ab initio* reference data consisted of relative conformational energies, nearly all of which were determined either from the composite "MP4SDQ/TZP" calculations carried out at MP2/6-31G*-optimized geometries for ~380 conformers (Set A) or from single-point MP2/TZP calculations carried out at ~1450 torsionally incremented geometries derived from MP2/6-31G*-optimized geometries (Set B). Benchmark calculations using still higher levels of theory and comparisons to experiment showed these to be the best tractable levels currently available to us.[26] Set A afforded 249 comparisons of "MP4SDQ/TZP" energies for optimized equilibrium or torsionally constrained conformers. Set B in turn yielded 1192 energy comparisons, each of which relates the MP2/TZP energy of a structure derived from a MP2/6-31G*-optimized equilibrium conformer to that of a "torsion profile" structure obtained by rotating one torsion bond by a specified extent (e.g., $\pm 30°, \pm 60°, \ldots$). The inclusion of these comparisons assured that MMFF94 has a reasonable understanding of torsional profiles and barriers. Full details are given in part IV.[26]

We view the determination of torsion parameters as a particularly strong component in the development of core MMFF94. No other force field, to our knowledge, has employed so broad a range of comparably accurate data on conformational energies in its derivation.

DETERMINATION OF MUTUALLY CONSISTENT MMFF PARAMETERS

Most force fields, have been derived using a "functional group" approach in which, for instance, "hydrocarbon parameters" are determined by fitting to data on alkanes and are then frozen. When alcohols and ethers, for example, are fit, only the parameters that arise from the newly introduced oxygen and polar hydrogen atom types need to be determined. This approach greatly sim-

plifies the derivation of the force field but fails to allow for the possibility that correlations between parameters may yield values that fit the limited original data (e.g., on hydrocarbons) well but are poorly defined and/or are not optimal for describing subsequent data (e.g., for hydrocarbon fragments in alcohols and ethers, etc.).

A better strategy would be to determine all the force-field parameters simultaneously from the full set of experimental and/or computational data. Such an approach would ensure that any shortcomings in the performance of the force field would be attributable to its form, or to the quality of the data used, rather than to its means of parameterization. This approach is computationally impractical at the present time. Fortunately, however, many classes of force-field parameters depend only weakly on others. For example, quadratic force constants change modestly when small changes are made in molecular geometries, and reference bond lengths and angles are insensitive to values employed for torsion parameters. This weak dependence allowed us to fashion a composite strategy that provided a computationally tractable approximation to the ideal of simultaneous determination of all parameters. We implemented this strategy by carrying out between three and four interactions over the set of procedures described in the previous three subsections; for good measure, we also redetermined the nonbonded parameters for hydrogen-bonding interactions, as described earlier, before the final determination of the torsion parameters. This approach allowed each class of parameters to be determined in the context of successively refined values for parameters belonging to other classes. As a result, nearly all parameters derived by this set of procedures have been determined in a physically self-consistent fashion.[59]

Performance of MM2X and MMFF94

This section summarizes MMFF94's ability to reproduce *ab initio* data used in its parameterization and also notes how well MM2X[46] performs. Further details may be found in the accompanying studies.[24-26]

MOLECULAR DIPOLE MOMENTS

For MMFF94, the partial atomic charges calculated from the computationally derived bond charge increments reproduced the set of 423 HF/6-31G* molecular dipole moments, increased by 10% as described above, with the rms deviations[60] shown:

	MMFF94	MM2X
Dipole magnitude	0.39 D	0.64 D
Dipole direction	5.5°	10.8°

Also listed are the results obtained using the less widely parameterized MM2X force field for a somewhat smaller set of HF/6-31G* dipole moments. For comparison, the rms value of the scaled HF/6-31G* dipole moments is 3.42 D. Thus, the average MMFF94 error is slightly larger than 10%, while that for MM2X is closer to 20%. MMFF94 is also considerably more accurate for dipole directions. The present performance, and that for intermolecular interaction energies and geometries in hydrogen-bonded dimers,[24] appears quite reasonable for an approach that is simple enough to allow virtually automatically application to a wide range of organic and bio-organic systems. Nevertheless, the treatment of electrostatic interactions is one area in which improvement particularly needs to be made in the future.

EQUILIBRIUM BOND LENGTHS

The comparisons shown below involved a total of 4205 equilibrium bond lengths. They were obtained by using MMFF to optimize 358 MP2/6-31G* equilibrium conformers and by systematically comparing the *ab initio*- and MMFF-optimized geometries.[25] For MM2X, 324 conformers covered by its parameterization were optimized, yielding a comparison of 3850 bond lengths. The results, cited below, are stated as rms deviations in angstroms from the optimized MP2/6-31G* bond lengths:

	MMFF94	MM2X
Equilibrium bond lengths	0.006	0.018

Clearly, the results for MMFF94 are excellent, those for MM2X respectable.

EQUILIBRIUM BOND ANGLES

A total of 7021 equilibrium bond angles for MMFF94 and 6462 bond angles for MM2X were examined. The results are stated as rms deviations

in degrees from the optimized MP2/6-31G* bond angles:

	MMFF94	MM2X
Equilibrium bond angles	1.16°	1.70°

Here, too, the MMFF94 results are quite good. MM2X also performs reasonably well.

WILSON OUT-OF-PLANE PUCKERING ANGLES

For MMFF94, 237 conformers had out-of-plane centers involving a total of 1926 out-of-plane angles. For MM2X, 206 conformers had 1755 such angles. Many of the comparisons are of little interest, however, as all the methods find carbonyl and olefinic carbon to be essentially planar in unstrained compounds. In contrast, nonplanarity at nitrogen is found in the MP2/6-31G* structures for aliphatic amines, for most amides (which in this work include hydroxamic acids), and for such "unsaturated" amines as amidines, guanidines, vinylic amines and aromatic amines. For these classes, the following MMFF94 and MM2X[61] rms deviations were found:

	MP2 rms angle	MMFF94 rms dev.
Amides (183 angles)	22.6°	9.38°
Unsaturated amines (33 angles)	43.8°	2.05°
Saturated amines (96 angles)	57.5°	0.91°
	MP2 rms angle	**MM2X rms dev.**
Amides (153 angles)	18.5°	17.5°
Unsaturated amines (33 angles)	43.8°	43.8°
Saturated amines (84 angles)	56.9°	2.24°

Shown for comparison are the rms values of the MP2/6-31G*-optimized Wilson angles.

Clearly, MMFF94 is far superior, though even it encounters some difficulty with amides, whose nitrogen center is notoriously easy to deform.[62] As we show in part III,[25] however, MMFF94 gives rms values for Wilson angles in primary, secondary, and tertiary amides that correctly reproduce the degree of puckering found in the MP2/6-31G* structures in an overall sense. MMFF94 also correctly describes the nonplanar equilibrium geometries of unsaturated amines, whereas MM2X does not.

TORSION ANGLES

Comparisons for a total of 7974 torsion angles for MMFF94 and 7409 for MM2X gave rms deviations from the MP2/6-31G*-optimized torsion angles as stated below:

	MMFF94	MM2X
Torsion angles	5.83°	11.38°

As discussed in part III,[25] MMFF94 sometimes underestimates and sometimes overestimates the degree of pyramidalization at nitrogen found in the MP2/6-31G* structures for amides. The associated errors in out-of-plane angles also affect the computed torsion angles and contribute significantly to the cited overall rms deviation. In addition, in a number of cases involving methyl groups attached to sp^2-hybridized centers, the relative energies for the torsionally incremented (Set B) structures on the MP2/TZP surface suggest that the MP2/6-31G* geometries (to which comparison is being made) are not equilibrium conformers on the higher level surface (from which the torsion parameters were largely derived). When questionable cases involving methyl rotations are excluded, the rms deviation for MMFF94 falls to about 5°.[26]

CONFORMATIONAL ENERGIES AND TORSION PROFILES

As previous described, the torsion parameters were derived in fits to two sets of data on conformational energies. The results, stated as rms deviations in kilocalories per mole, were as follows:

	MMFF94	MM2X
Conformational energies (Set A)	0.31	1.12
Torsion-profile energies (Set B)	0.50	1.57

For comparison, rms values for the relative energies were 3.88 kcal/mol for the conformational energies and 4.37 kcal/mol for the torsion-profile energies for MMFF94, and 2.30 and 4.38 kcal/mol, respectively, for MM2X.[63] Thus, MMFF94 accounts for about 90% of the variation in the *ab initio*

relative energies in each case, whereas MM2X accounts for about 50%.

We note that Sets A and B each contained extensive conformational data on the glycine and alanine dipeptide analogs[64] and on the full, methyl-capped glycine and alanine dipeptides. As we show in part IV,[26] MMFF94 reproduces these data very well. All common protein sidechains are also covered in its parameterization. No other published force field, to our knowledge, has been derived from a comparably extensive set of high-quality data on conformational comparisons pertinent to simulations on proteins.

ADDITIONAL COMPARISONS

For the linear water dimer (optimized with the O—H ··· O angle restricted to 180°), MMFF94 gives a dimerization energy of −6.53 kcal/mol, an O ··· O distance of 2.75 Å, and an angle between the O ··· O axis and the acceptor H—O—H plane of 27°.[24] The analogous quantities[65] are −6.50 kcal/mol, 2.74 Å and 27° for TIP3P water; −6.59 kcal/mol, 2.75 Å and 26° and for SPC[37] water; and −6.24 kcal/mol, 2.74 Å and 46° and for TIP4P water. Thus, "MMFF water" behaves similarly in this static test. Work using liquid-phase simulations is currently underway to test and, if necessary, to reformulate or reparameterize MMFF94.[66]

Results for geometries and interaction energies for an extensive series of hydrogen-bonded dimers are presented in part II.[24] The comparisons show that MMFF94 accurately reflects the trends in interaction energies and geometries manifested in the *ab initio* calculations. The force field therefore appears to properly balance the strengths of water–water, water–solute, and solute–solute interactions. These comparisons suggest that MMFF94 can be used with confidence in computational studies of ligand-receptor binding. Also given in part II[24] are comparisons of vdW interaction energies for the $(CH_4)_2$ and $(H_2)_2$ homodimers as a function of separation and orientation. These comparisons show that MMFF94 accounts reasonably well for prototype nonpolar vdW interactions.

Accuracy in Predicting Experiment

This section summarizes MMFF94's ability to reproduce experimental data. Further details may be found in parts III and IV[25,26]; comparisons to experiment for the extended MMFF94 parameterization are given in part V.[27]

MOLECULAR GEOMETRIES

We have compared MMFF94 to experiment and to published MM3 geometries for a series of 30 organic molecules covering a variety of functional groups. For bond lengths, rms deviations relative to experiment were found to be 0.014 Å for MMFF94 and 0.010 Å for MM3; for bond angles, the rms deviation was 1.2° for each force field. Thus, MMFF94 is as successful as MM3 in predicting experimental bond angles, despite the fact that no experimental data on molecular geometries was used in deriving MMFF94. MM3 predicts experimental bond lengths more accurately, even in this test in which the experimental bond lengths were not strictly limited to the r_g values MM3 seeks to emulate, but whether a difference of this magnitude is of practical significance for molecular simulations is unknown. Some of the difference in predicting experimental bond lengths arises from small, systematic deviations from experiment in the underlying MP2/6-31G* bond lengths. Part arises from the intrinsic difference between energy-minimized (MP2/6-31G*) and thermally averaged (experimental) bond lengths; as a force field intended for use in molecular-dynamics simulations, MMFF94 reproduces the energy-minimum bonds lengths obtained from the *ab initio* calculations, whereas MM3 incorporates thermal-averaging effects into its static model. For torsion angles, one particularly notable difference occurs for the "cisoid" conformation of 1,3-butadiene, for which MMFF94 predicts a nonplanar energy-minimized structure, wheeas MM3 gives a planar structure. A more complete discussion, including comparisons to CHARMm[9] and UFF[11] and more detailed comparisons to MM3, is given in part III.[25]

VIBRATIONAL FREQUENCIES

To further characterize MMFF94's performance, vibrational frequencies were calculated for formamide, benzene, formic acid, formaldehyde, acetaldehyde, methylamine, ammonia, methanol, water, methane, ethane, ethylene, hydrogen sulfide, *gauche*-ethanethiol, and dimethyl disulfide. When compared with published MM3 and experimentally determined frequencies,[25] rms deviations versus experiment were found to be 61 cm^{-1} for MMFF94, 57 cm^{-1} for MM3 for the slightly smaller

subset of molecules for which MM3 vibrational frequencies had been published, and 60 cm^{-1} for MMFF94 for the same set of molecules used in assessing MM3. Thus, MMFF94 and MM3 perform comparably on an overall basis. For MM3, however, we noted a number of instances in which its parameterization had employed experimental frequencies that differed significantly—by nearly 400 cm^{-1} for a B_{2u} stretching mode in benzene—from other published experimental values that themselves had been shown to be compatible with theoretically calculated frequencies. Such instances illustrate one of the hazards of deriving a force field from experimental data: such data, at times, can contain large errors that then become a part of the derived force field.[25]

COFORMATIONAL ENERGIES AND ROTATIONAL BARRIERS

Energy differences calculated using MMFF94 reproduce a diverse set of 37 experimentally determined gas-phase and solution conformational energies, enthalpies, and free energies (rms value 2.3 kcal/mol), with an rms deviation of 0.38 kcal/mol, as opposed to 0.37 kcal/mol for both the supporting "MP4SDQ/TZP" calculations and for MM3.[26] Moreover, MMFF94 reproduces 28 experimentally determined rotational barriers (rms value 3.7 kcal/ mol) with a rms deviation of 0.39 kcal/mol. Importantly, these comparisons, and others discussed in part IV,[26] demonstrate that fitting MMFF94 to high-quality theoretical data has simultaneously conferred the ability to fit experiment. MMFF94 can be expected to perform equally well for the wide range of systems for which it has been parameterized but for which little or no experimental data are available.

Implementation of MMFF in OPTIMOL, CHARMm, and BatchMin

In this section, we discuss some pertinent elements related to the implementation of MMFF94 in OPTIMOL,[46] the host molecular-mechanics platform for which MMFF94 and MM2X were developed; the same elements also apply to the recent implementations of MMFF93 in CHARMm[28] and of MMFF94 in BatchMin.[29]

In each of these implementations, the user (or the invoking modeling platform) simply represents the subject molecule in language familiar to the organic chemist, that is, as a collection of atoms joined by single, double, or triple bonds, some atoms of which may have a nonzero formal charge. Aromatic systems may be supplied in any constituent Kekule form. The program then uses the supplied structural information to generate all additional information needed to carry out the calculation. It automatically determines the torsional "tree structure," perceives and classifies rings, defines symbolic atom types based on local connectivity, detects aromaticity, and creates appropriate lists of bond, angle, and torsional interactions. As previously described,[39] the symbolic atom types (cf. Table III) are then translated into the numeric values used to assign force-field parameters to the force-field interaction terms.

In establishing the relationship between parameters and force-field interactions, the parameter files, which are kept in "canonical order" based on indices derived from the numerical atom types, are processed using a rapid binary search algorithm. If present, the *fully qualified* parameter corresponding to the precise set of atom types (supplemented, in ambiguous situations, by defined bond, angle, stretch–bend, or torsion "interaction types"[25,26]) is retrieved and used. For vdW, bond stretching, stretch–bend, and bond-charge-increment parameters, no *equivalences* are recognized. For angle bending, out-of-plane bending, and torsion interactions, however, MMFF94 executes a staged "step-down" procedure in which increasingly generic values are sought whenever the "fully qualified" parameter is not found. This protocol is governed by the entries in Table IV, where the "Level 1" atom types define the fully qualified parameters. Entries from Levels 2–5 are employed as needed in subsequent searches; those at Level 5, always "0" except for atomic ions, serve as *wild cards*. Such wild card values are used only for peripheral atoms in an angle bending or torsional interaction or for noncentral atoms in an out-of-plane interaction. Level 4 generally corresponds to the atomic species, and Level 3 to atomic species plus hybridization. Currently, the first two levels employ identical numerical atom types. Unique values for Level 1 may be specified later if certain atom types need to be defined more specifically.[67] The protocol used in the step-down procedure depends on the type of interaction (angle, torsion, out-of-plane).[68] If no parameter is found, one of a series of carefully calibrated empirical rules is invoked (cf. part V[27]). This staged-search/default-rule procedure allows applications to go forward when specific parameters are un-

TABLE IV.
Numerical Atom Type Equivalences Used in Assigning MMFF94 Parameters

MMFF symbol[a]	Equivalence level[b] 1	2	3	4	5
CR	1	1	1	1	0
C=C	2	2	2	1	0
C=O	3	3	3	1	0
CSP	4	4	4	1	0
HC	5	5	5	5	0
OR	6	6	6	6	0
O=C	7	7	7	6	0
NR	8	8	8	8	0
N=C	9	9	9	8	0
NC=O	10	10	10	8	0
F	11	11	11	11	0
CL	12	12	12	12	0
BR	13	13	13	13	0
I	14	14	14	14	0
S	15	15	15	15	0
S=C	16	16	16	15	0
SO	17	17	17	15	0
SO2	18	18	18	15	0
SI	19	19	19	19	0
CR4R	20	20	1	1	0
HOR	21	21	21	5	0
CR3R	22	22	22	1	0
HNR	23	23	23	5	0
HOCO	24	24	24	5	0
PO4	25	25	25	25	0
P	26	26	26	25	0
HN=C	27	27	28	5	0
HNCO	28	28	28	5	0
HOCC	29	29	29	5	0
CE4R	30	30	2	1	0
HOH	31	31	31	31	0
O2CM	32	32	7	6	0
HOS	33	33	21	5	0
NR+	34	34	8	8	0
OM	35	35	6	6	0
HNR+	36	36	36	5	0
CB	37	37	2	1	0
NPYD	38	38	9	8	0
NPYL	39	39	10	8	0
NC=C	40	40	10	8	0
CO2M	41	41	3	1	0
NSP	42	42	42	8	0
NSO2	43	43	10	8	0
STHI	44	44	16	15	0
NO2	45	45	10	8	0
N=O	46	46	9	8	0
NAZT	47	47	42	8	0
NSO	48	48	9	8	0
O+	49	49	6	6	0
HO+	50	50	21	5	0
O=+	51	51	7	6	0

TABLE IV. (continued)

MMFF symbol[a]	Equivalence level[b] 1	2	3	4	5
HO=+	52	52	21	5	0
=N=	53	53	42	8	0
N+=C	54	54	9	8	0
NCN+	55	55	10	8	0
NGD+	56	56	10	8	0
CGD+	57	57	2	1	0
NPD+	58	58	10	8	0
OFUR	59	59	6	6	0
C%	60	60	4	1	0
NR%	61	61	42	8	0
NM	62	62	10	8	0
C5A	63	63	2	1	0
C5B	64	64	2	1	0
N5A	65	65	9	8	0
N5B	66	66	9	8	0
N2OX	67	67	9	8	0
N3OX	68	68	8	8	0
NPOX	69	69	9	8	0
OH2	70	70	70	70	70
HS	71	71	5	5	0
S2CM	72	72	16	15	0
SO2M	73	73	18	15	0
=S=O	74	74	17	15	0
-P=C	75	75	26	25	0
NM5	76	76	9	8	0
CLO4	77	77	12	12	0
C5	78	78	2	1	0
N5	79	79	9	8	0
CIM+	80	80	2	1	0
NIM+	81	81	10	8	0
N5AX	82	82	9	8	0
FE+2	87	87	87	87	87
FE+3	88	88	88	88	88
F-	89	89	89	89	89
Cl-	90	90	90	90	90
Br-	91	91	91	91	91
LI+	92	92	92	92	92
K+	93	93	93	93	93
NA+	94	94	94	94	94
ZN+2	95	95	95	95	95
CA+2	96	96	96	96	96
CU+1	97	97	97	97	97
CU+2	98	98	98	98	98
MG+2	99	99	99	99	99

[a] Shown are representative MMFF94 symbolic atom types (cf. Table III).

[b] The Level 1 numerical atom types are the primary values. The usage of the equivalences reflected in Levels 2–5 is described in the "Implementation of MMFF94 in OPTIMOL, CHARMm, and BatchMin" section.

available, though inevitably with a loss in reliability.

Concluding Discussion

This and the accompanyig studies[24-27] introduce MMFF94, the initial published version of the Merck Molecular Force Field. As was noted in the Introduction, this version of MMFF is primarily intended for use in molecular-dynamics studies; a modified version intended for use in energy-minimization studies is under development.[32]

MMFF94's formulation and parameterization has a number of distinguishing features. One is that MMFF94 uses a unique functional form for describing van der Waals interactions and employs novel combination rules that embody a systematic correlation of vdW parameters with those that describe experimentally well-characterized interactions involving small molecules and rare-gas atoms.[39]

A second distinguishing feature is that MMFF94's core parameterization is primarily based on a large amount of computational data obtained from *ab initio* calculations—approximately 500 molecular structures optimized at the HF/6-31G* level, 475 structures optimized at the MP2/6-31G* level, 380 structures evaluated at the composite "MP4SDQ/TZP" level[26] using MP2/6-31G*-optimized geometries, and 1450 structures evaluated in single-point calculations at the MP2/TZP level. While *ab initio* data have been used in force development for at least two decades,[69] no other effort, to our knowledge, has used so much data of such high quality.

A third distinguishing feature is that the core, computationally derived, portion of MMFF94 has been parameterized for an unusually wide variety of chemical systems. As a result, MMFF94 provides well-defined parameters for more than 20 chemical families and treats many frequently occurring combinations of functional groups. The range of coverage for the extended parameterization is far larger still.[27]

The methodology used in parameterizing MMFF94 represents a fourth distinguishing feature. Specifically, nearly all MMFF94 parameters have been determined in a mutually consistent fashion[59] from the full set of available computational data. Other force-field derivations have usually employed a "functional group" approach in which certain parameters are fit to a portion of the available data and are then frozen. While practical limitations of the "functional group" approach have not yet been convincingly demonstrated, we prefer an approach that, by construction, yields mutually consistent values for the parameters.

These attributes of its functional form and parameterization combine to produce a force field that, by contemporary standards, performs very well. In particular, both computational data and experimental data are described well—the latter to a degree comparable to that achieved by MM3. These comparisons demonstrate that MMFF94's parameterization against computational data has simultaneously conferred the ability to reproduce experiment. Consequently, MMFF94 can be expected to perform equally well throughout the range of its parameterization from high-quality computational data, even for the many systems for which relevant experimental data is unavailable. This attribute constitutes a particularly strong advantage of a computationally derived force field like MMFF94. Comparisons in functional form, performance, and/or manner of derivation for such other force fields as MM3, CFF93, OPLS, AMBER, CHARMm, UFF, and DREIDING are given in the accompanying studies.[24-27]

While the computational data used in its derivation necessarily relate to small molecules, it should be emphasized that MMFF94 has consciously been designed to be both a "small molecule" and a "protein" force field. Among other factors, the excellent results obtained for conformational energies for dipeptide analogs and for dipeptides[26] and the uniform and balanced parameterization that has been pursued for nonbonded solvent–solvent, solvent–solute, and solute–solute interactions,[24] when taken together with the reproduction of experimental data for small molecules with an accuracy comparable to that of MM3, suggest that MMFF94 should perform well in both domains.

Despite encouraging success, certain limitations are evident. One of particular importance arises from the fact that MMFF94 uses static atom-centered charges. As such, it neglects both higher order multipoles and electrostatic effects that arise from molecular polarizability. Because of these simplifications, MMFF94, like a number of other force fields, employs "enhanced" charge distributions that emulate the effect of polarizability in amplifying electrostatic interactions for favorable contacts in a high-dielectric medium. Unfortunately, these enhanced charge distributions also amplify electrostatically unfavorable iteractions,

whereas proper account of polarizability would diminish them. They also improperly enhance electrostatic interactions in gas-phase or low-dielectric environments. Furthermore, they may not be optimal for describing *intramolecular* interactions, and may thereby limit the ability of the force field to account for differences in conformational energies. Indeed, compounds containing two or more strongly polar functional groups in close proximity have proven to be the most problematic in this respect, though good results have been obtained in most cases to date.[26] Other significant limitations include: the overly simplistic nature of the bond-charge-increment scheme used to assemble the partial atomic charges[24]; the lack of conformational dependence of the resultant charges[24]; and the omission of bond–torsion (and certain other) cross terms needed to account for significant geometrical changes that can occur when a torsion angle varies,[25] an example being the elongation of an amide partial C—N double bond by up to 0.1 Å when conjugation is broken. A further significant limitation is that no account is taken of metal–ligand interactions beyond that afforded by a relatively simplistic model that includes only electrostatic and van der Waals nonbonded interactions.

What can be expected from future efforts at force-field development? First and foremost, better physical forms will need to be employed, particularly for electrostatic interactions.[70] For example, even highly regarded water models such as SPC and TIP3P are known to describe certain configurations for the water dimer very poorly.[71] In addition, a broader selection of cross terms than are employed in MMFF94 will almost certainly be needed, and other enhancements can also be envisioned.[24–27] We expect that a computational approach based almost solely on the use of *ab initio* data will become indispensable and that reliance on experimental data will diminish. The problem, ultimately, is one of information: too many force-field parameters, too little experimental data, and in many instances too nebulous a relationship between the two. Fortunately, significant improvements in computer technology can be expected to make it increasingly feasible both to use more complex force fields in molecular simulations and to employ ever more rigorous computational models to generate the data needed to derive them. But of course this approach will still yield a gas-phase force field, whereas most applications of interest to pharmaceutical and medicinal chemists take place in the condensed phase. This observation brings us back to an objective that underlies this work but has not yet been clearly stated: *to define a force field that describes gas-phase molecular properties accurately and that behaves properly when the gas-phase system is embedded in the condensed phase.* This objective cannot fully be met in a force field that treats electrostatic interactions as simplistically as does the present version of MMFF. Ultimately, however, it will be met, because "only" physics is involved, and because that physics is becoming increasingly well understood.[20,21]

Acknowledgment

The author thanks the many colleagues at the Merck Research Laboratories who have supported this work; have contributed to the development of OPTIMOL, in which MMFF94 was first implemented; have developed computational tools used in MMFF94's derivation; have created and maintained the computational environment at Merck; or have utilized MMFF94 and its predecessor force fields in their own research. Many of these contributions are explicitly cited in the accompanying articles.[24–27]

Supplementary Material

Appendix A (definition and role of the 16 MMFF94 parameter files)[72] and Appendix B (computer-readable ASCII file containing the MMFF94 parameter files[23]) are available in Supplementary Material.

References

1. J. B. Hendrickson, *J. Am. Chem. Soc.*, **83**, 4537–4547 (1961). For an even earlier reference, see: F. H. Westheimer, in *Steric Effect in Organic Chemistry*, M. S. Newman, Ed., Wiley, New York, 1956, Chapter 12.
2. D. H. Wertz and N. L. Allinger, *Tetrahedron*, **30**, 1579 (1974).
3. (a) N. L. Allinger, *J. Am. Chem. Soc.*, **89**, 8127 (1977); (b) U. Burkert and N. L. Allinger, *Molecular Mechanics*; American Chemical Society, Washington, DC, 1982; (c) N. L. Allinger and Y. Yuh, *QCPE*, **12**, 395 (1980).
4. N. L. Allinger, Y. H. Yuh, and J-H. Lii, *J. Am. Chem. Soc.*, **111**, 8551–8566 (1989). See also: N. L. Allinger and L. Yan, *J. Am. Chem. Soc.*, **115**, 11918–11925 (1993), and references therein.
5. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.*, **117**, 5179–5197 (1995); S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comput. Chem.*, **7**, 230–252 (1986); S. J.

Weiner, P. A. Kollman, D. T. Nguyen, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta Jr., and P. Weiner, *J. Am. Chem. Soc.*, **106**, 765–784 (1984).

6. (a) W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, **110**, 1657–1666 (1988), and references therein; (b) H. A. Carlson, T. B. Nguyen, M. Orozco, and W. J. Jorgensen, *J. Comput. Chem.*, **14**, 1240–1249 (1993), and references therein.
7. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.*, **4**, 187–217 (1983).
8. S. Lifson, A. T. Hagler, and P. Dauber, *J. Am. Chem. Soc.*, **101**, 5111–5121 (1979), and references therein.
9. F. M. Momany and R. Rone, *J. Comput. Chem.*, **13**, 888–900 (1992).
10. S. L. Mayo, B. D. Olafson, and W. A. Goddard III, *J. Phys. Chem.*, **94**, 8897 (1990).
11. A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard III, and W. M. Skiff, *J. Am. Chem. Soc.*, **114**, 10024–10035 (1992), and references therein.
12. A. Vedani and D. W. Huhta, *J. Am. Chem. Soc.*, **112**, 4759–4767 (1990).
13. V. S. Allured, C. M. Kelly, and C. R. Landis, *J. Am. Chem. Soc.*, **113**, 1 (1991).
14. D. M. Root, C. R. Landis, and T. Cleveland, *J. Am. Chem. Soc.*, **115**, 4201–4209 (1993).
15. J. R. Maple, U. Dinur, and A. T. Hagler, *Proc. Natl. Acad. Sci. USA*, **85**, 5350–5354 (1988). A. T. Hagler, J. R. Maple, T. S. Thacher, G. B. Fitzgerald, and U. Dinur, In *Computer Simulation of Biomolecular Systems*, W. F. van Gunsteren and P. K. Weiner, Eds., ESCOM, Leiden, 1989, pp. 149–167. U. Dinur and A. T. Hagler, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1991, Vol. 2, pp. 99–163. See also: J. Palca, *Nature*, **322**, 586 (1986).
16. (a) J. R. Maple, M.-J. Hwang, T. P. Stockfish, U. Dinur, M. Waldman, C. S. Ewig, and A. T. Hagler, *J. Comput. Chem.*, **15**, 161–182 (1994); M.-J. Hwang, T. P. Stockfish, and A. T. Hagler, *J. Am. Chem. Soc.*, **116**, 2515–2525 (1994).
17. N. L. Allinger, K. Chen, and J.-H. Lii, *J. Comput. Chem.* (this issue).
18. (a) W. J. Hehre, L. Radom, P. v. R Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory*, Wiley, New York, 1986, Chapter 6; (b) D. J. DeFrees, B. A. Levi, S. K. Pollack, W. J. Hehre, J. S. Binkley, and J. A. Pople, *J. Am. Chem. Soc.*, **101**, 4085–4089 (1979).
19. For a recent example of a significant experimental error which was corrected on the basis of information from *ab initio* calculations, including calculations at the MP2/6-31G* level used in the present work, see: B. J. Smith and L. Radom, *J. Am. Chem. Soc.*, **112**, 7525–7528 (1990).
20. See, for example: (a) U. Dinur and A. T. Hagler, *J. Chem. Phys.*, **91**, 2949–2958 (1989); (b) U. Dinur and A. T. Hagler, *J. Chem. Phys.*, **91**, 2959–2970 (1989); (c) U. Dinur, *J. Comput. Chem.*, **12**, 91–105 (1991); (d) U. Dinur, *J. Comput. Chem.*, **12**, 469–486 (1991); (e) U. Dinur, *J. Phys. Chem.*, **94**, 5669–5671 (1990).
21. See, for example: (a) G. Corongiu, M. Migliore, and E. Clementi, *J. Chem. Phys.*, **90**, 4629 (1989); (b) L. X. Dang, J. E. Rice, J. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.*, **113**, 2481–2486 (1991), and references therein; (c) M. Sprik, *J. Phys. Chem.*, **95**, 2283–2291 (1991), and references therein. (d) S.-B. Zhu, S. Yao, J.-B. Zhu, S. Singh, and G. W. Robinson, *J. Phys. Chem.*, **95**, 6211–6217 (1991); (e) S. W. Rick, S. J. Stuart, and B. J. Berne, *J. Chem. Phys.*, **101**, 6141–6156 (1994); (f) D. N. Bernardo, Y. Ding, K. Krough-Jespersen, and R. M. Levy, *J. Phys. Chem.*, **98**, 4180–4187 (1994); (g) For an early implementation of induced-dipole effects, see also L. Dosen-Micovic, D. Jeremic, and N. L. Allinger, *J. Am. Chem. Soc.*, **105**, 1716, 1723 (1983).
22. S. Dasgupta and W. A. Goddard III, *J. Chem. Phys.*, **90**, 7207–7215 (1989).
23. The MMFF94 parameters (Appendix B, Supplementary Material) are available in computer-readable form (see footnote * on first page of this article).
24. Part II: T. A. Halgren, *J. Comput. Chem.* (this issue).
25. Part III: T. A. Halgren, *J. Comput. Chem.* (this issue).
26. Part IV: T. A. Halgren and R. B. Nachbar, *J. Comput. Chem.* (this issue).
27. Part V: T. A. Halgren, *J. Comput. Chem.* (this issue).
28. This collaboration involved Prof. Martin Karplus (Harvard University) and Dr. Ryszard Czerminski and others of Molecular Simulations, Inc. (San Diego, CA). Currently, a version of CHARMm that supports the earlier and less widely parameterized MMFF93 force field (which lacks, e.g., the ability to recognize a number of the ionic species parameterized in ref. 27; see also refs. 25 and 26) is available from MSI. However, while the local Merck code for CHARMm employs MMFF94, arrangements for including MMFF94 in the distributed MSI version have not yet been concluded.
29. P. S. Shenkin and T. A. Halgren (work in progress). The MacroModel program suite and its BatchMin module, developed in the laboratories of Professor Clark Still, are available from Columbia University (New York, NY).
30. OPTIMOL has been developed and maintained by the author, but is based in part on computer code adapted from a public domain version of MM2 or written by Drs. R. B. Nachbar, B. L. Bush, G. M. Smith, E. F. Fluder Jr., and J. D. Andose of the Merck Research Laboratories.
31. Distribution of OPTIMOL by the Quantum Chemistry Program Exchange (Indiana University) would permit free usage of the program but would prohibit its commercialization.
32. T. A. Halgren and R. B. Nachbar (work in progress).
33. *The Merck Index*, 11th ed., S. Budavari, Ed., Merck & Co., Rahway, NJ, 1989.
34. *Fine Chemicals Directory Handbook*, Fraser Williams (Scientific Systems), London, 1983–1985. Connection tables distributed by Molecular Design Ltd., Hayward, CA.
35. PROBE is a computer program used to derive molecular-mechanics parameters in least-squares fits to data obtained from *ab initio* calculations. PROBE was created for the Biosym Consortium on Potential Energy Functions by Biosym Technologies, Inc. (now Molecular Simulations, Inc.); cf. refs 15 and 16a. The derivation of MMFF94 used a 1991 version of PROBE.
36. See, for example, G. C. Lie and E. Clementi, *Phys. Rev.*, **33A**, 2679 (1986).
37. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, In *Intermolecular Forces*, B. Pullman, Ed., Reidel, Dordrecht, Holland, 1981, pp. 331–342.

38. B. M. Pettit, in *Computer Simulation of Biomolecular Systems*, W. F. van Gunsteren and P. K. Weiner, Eds., ESCOM, Leiden, 1989, pp. 94–100.
39. T. A. Halgren, *J. Am. Chem. Soc.*, **114**, 7827–7843 (1992).
40. W. J. Hehre, L. Radom, P. v. R Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory*, Wiley, New York, 1986, Chapter 4. The 6-31G* basis sets also known as 6-31G(d).
41. A. D. MacKerell Jr. and M. Karplus, *J. Phys. Chem.*, **95**, 10559–10560 (1991); A. D. MacKerell Jr., J. Wiórkiewicz-Kuczera, and M. Karplus, *J. Am. Chem. Soc.* (in press).
42. M. J. Frisch. J. E. Del Bene, J. S. Binkley, and H. F. Schaeffer III, *J. Chem. Phys.*, **84**, 2279–2289 (1986).
43. T. A. Halgren, *J. Am. Chem. Soc.*, **112**, 4710–4723 (1990).
44. M. Orozco and F. J. Luque, *J. Comput. Chem.*, **14**, 881–894 (1993).
45. E. B. Wilson Jr., J. C. Decius, and P. C. Cross, *Molecular Vibrations*, Dover, New York, 1955, Chapter 4.
46. M. K. Holloway, J. M. Wai, T. A. Halgren, P. M. D. Fitzgerald, J. P. Vacca, B. D. Dorsey, R. B. Levin, W. J. Thompson, L. J. Chen, S. J. deSolms, N. Gaffin, A. K. Ghosh, E. A. Giuliani, S. L. Graham, J. P. Guare, R. W. Hungate, T. A. Lyle, W. M. Sanders, T. J. Tucker, M. Wiggins, C. M. Wiscount, O. W. Woltersdorf, S. D. Young, P. L. Darke, and J. A. Zugay, *J. Med. Chem.*, **38**, 305–317 (1995).
47. Note that this truncation does not lead to a "cubic-bend" catastrophe because the range of the angle is limited to 180° and the cubic-bend constant is relatively small.
48. See, for example, the "vdW and hydrogen bonding" parameters listed in Table I of: J.-H. Lii and N. L. Allinger, *J. Comput. Chem.*, **12**, 186–199 (1991).
49. M. Waldman and A. T. Hagler, *J. Comput. Chem.*, **14**, 1077–1084 (1993).
50. Most of the *ab initio* calculations used in parameterizing MMFF were performed on the Merck Research Laboratories Cray YPM8i/4-128 supercomputer.
51. M. J. Frisch, M. Head-Gordon, H. B. Schlegel, K. Raghavachari, J. S. Binkley, C. Gonzalez, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, R. L. Martin, L. R. Kahn, J. J. P. Stewart, E. M. Fluder, S. Topiol, and J. A. Pople, Gaussian 88, Gaussian, Inc., Pittsburgh, PA, 1988, as modified at Merck for improved I/O performance by E. M. Fluder.
52. M. J. Frisch, M. Head-Gordon, G. W. Trucks, J. B. Foresman, H. B. Schlegel, K. Raghavachari, M. Robb, J. S. Binkley, C. Gonzalez, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, R. L. Martin, L. R. Kahn, J. J. P. Stewart, S. Topiol, and J. A. Pople, GAUSSIAN 90 (Revision J), Gaussian, Inc., Pittsburgh, PA, 1990.
53. M. L. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople, GAUSSIAN 92 (Revision C), Gaussian, Inc., Pittsburgh, PA, 1992.
54. Calculations at the MP2/6-31G*//MP2/6-31G* level are used as the basis of structure determination in the high-level composite G1 and G2 methods developed by Pople and coworkers; cf. L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, *J. Chem. Phys.*, **94**, 7221–7230 (1991), and references therein. Comparisons to experiment for a number of the molecules employed in the derivation of MMFF94 are summarized in A. St.-Amant, W. D. Cornell, T. A. Halgren, and P. A. Kollman, *J. Comput. Chem.*, **16**, 1483–1506 (1995).
55. See, for example, G. Chalasinski, M. M. Szczesniak, P. Cieplak,, and S. Scheiner, *J. Chem. Phys.*, **94**, 2873–2883 (1991), and references therein.
56. The ESP-fit calculations were carried out with a version of Gaussian 88 to which Dr. M. D. Miller (Merck Research Laboratories) had interfaced PDM88 (D. E. Williams, *QCPE*, Program No. 568, 1988).
57. In these MMFF optimizations, weak penalty-function restraints were applied to the torsion angles to insure that comparable MMFF and *ab initio* conformations were being compared (cf. ref. 26).
58. TORFIT is a versatile program developed at the Merck Research Laboratories which derives torsional parameters via least-squares fits to relative conformational energies (cf. ref. 26).
59. The procedure used is not strictly "mathematically" self-consistent, however, because formal couplings between parameters belonging to different classes (e.g., between reference values and force constants for angles at trigonal centers) have not been addressed. Further iterations would probably cause a slow drift away from the parameter values reported in this work. We view the parameters as being "physically" self-consistent, however, in the sense that such further iterations would not materially improve the fit to the computational data.
60. The cited rms deviations in dipole directions are weighted rms deviations constructed to avoid overemphasizing large errors in directions for dipole moments of small magnitude (cf. ref. 24).
61. We should note that MM2X actually uses the Allinger (MM2/MM3) definition for out-of-plane angles. To clarify the comparison to MMFF, however, we have used the Wilson definition in analyzing the MM2X-optimized geometries. The Allinger angles typically are about three times smaller in magnitude.
62. See, for example: M. W. Wong and K. B. Wiberg, *J. Phys. Chem.*, **96**, 668–671 (1992).
63. The first rms value is much lower for MM2X because some high-energy structures in the transition state region for C—N amide bond rotation in *N*-methylformamine were poorly treated by MM2X and had to be removed from the test set (cf. ref. 26).
64. For structures, see T. Head-Gordon, M. Head-Gordon, M. J. Frisch, C. L. Brooks III, and J. A. Pople, *J. Am. Chem. Soc.*, **113**, 5989–5997 (1991).
65. W. L. Jorgensen, J. Chandrasekhar and J. D. Madura, *J. Chem. Phys.*, **79**, 926–935 (1983).
66. (a) B. L. Bush, J. L. Banks, R. Czerminski, and T. A. Halgren (work in progress), using a local version of CHARMm into which MMFF94 has been integrated; (b) T. A. Halgren, S.-S. So, and M. Karplus (work in progress).
67. As an example, we might at some point wish to define different bond-charge increments for C=O groups in amides, esters, ketones, etc., for which differing symbolic atom types but common numeric atom types currently are assigned. The equivalence procedure provides a convenient way to do so without requiring that atom types and parameters describing common bond, angle, torsion, and other interactions simultaneously be modified.

68. For bending of the i–j–k angle, a five-stage process based in the level combinations 1–1–1, 2–2–2, 3–2–3, 4–2–4, and 5–2–5 is used. For i–j–k–l torsion interactions, a five-stage process based on level combinations 1–1–1–1, 2–2–2–2, 3–2–2–5, 5–2–2–3, and 5–2–2–5 is used, where stages 3 and 4 correspond to "half-default" or "half-wild-card" entries. For out-of-plane bending $ijk; l$, where j is the central atom [cf. eq. (5)], the five-stage protocol 1–1–1; 1, 2–2–2; 2, 3–2–3; 3, 4–2–4; 4, 5–2–5; 5 is used. The final stage provides wild-card defaults for all except the central atom.

69. N. L. Allinger and M. J. Hickey, *Tetrahedron*, **28**, 2157–2161 (1972). Although not specifically referenced, the results of this work were used in developing a force field for carbonyl compounds: N. L. Allinger, M. T. Trible, and M. A. Miller, *Tetrahedron*, **28**, 1173–1190 (1972) (N. L. Allinger, private communication).

70. W. F. van Gunsteren, In *Computer Simulation of Biomolecular Systems*, W. F. van Gunsteren and P. K. Weiner, Eds., ESCOM, Leiden, 1989, pp. 27–59.

71. M. Mezei and J. J. Dannenberg, *J. Phys. Chem.*, **92**, 5860–5861 (1988).

72. Appendix A is available in Supplementary Material (see footnote * on the first page of this article).